



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**Improving our understanding of the in vivo
modelling of psychotic disorders: a
systematic review and meta-analysis**

Zsanett Bahor

PhD Thesis

The University of Edinburgh

2018

“What gets us into trouble is not what we don't know.

It's what we know for sure that just ain't so.”

Mark Twain

Table of Contents

Declaration	i
Acknowledgements.....	iii
Publications and Conference Participation.....	v
Publications.....	v
Presentations	vi
List of Abbreviations	vii
Abstract	viii
Lay Summary	ix
1 Introduction.....	1
1.1 The Burden of Mental Health	1
1.2 Psychosis and Psychotic Disorders	2
1.2.1 Nosology	2
1.2.2 What we think we know about schizophrenia.....	4
1.2.3 Clinical management.....	16
1.2.4 Challenges in developing new treatments for psychotic disorders	17
1.2.5 Pre-clinical studies and the role of animal models	18
1.2.6 Translation of knowledge from basic science to the clinic and its failure	20
1.2.7 Systematic reviews and their utility	22
1.3 Aim and Objectives.....	24
2 Methods.....	27
2.1 Systematic Review	27
2.1.1 Stages of study selection and data extraction.....	31
2.1.2 Phase I: Pre-screening based on title and abstract.....	31
2.1.3 Phase II. Full text screening, categorization and assessment of reporting of experimental risk of bias	34
2.1.4 Phase III: Collection and Extraction of Data from Publications.....	38

2.2	Meta-analysis	45
2.2.1	Identification and definition of comparisons	45
2.2.2	Individual comparison effect size estimate	46
2.2.3	Weighted and pooling of effect sizes to give summary effect size	50
2.2.4	Assessing for heterogeneity	53
2.2.5	Exploring sources of heterogeneity	54
2.2.6	Bonferroni correction	56
2.3	Publication Bias.....	57
2.3.1	Funnel plotting.....	57
2.3.2	Egger regression.....	58
2.3.3	Trim and fill	58
3	Systematic Search Results and Characterization of the Literature	59
3.1	Introduction	59
3.2	Methods	59
3.3	Results	60
3.3.1	Phase I. Search of the literature	60
3.3.2	Phase II. Categorisation of studies of relevance.....	62
3.3.3	Outcome measures reported in the literature	68
3.3.4	Treatments reported in the literature	73
3.4	Discussion.....	76
3.4.1	Methods used to induce the model.....	76
3.4.2	Outcome measures reported.....	82
3.4.3	Therapeutic compounds reported to be tested	87
3.4.4	Limitations.....	91
4	Construct, Face, Predictive and External Validity: Developmental Animal Models of Schizophrenia	93
4.1	Introduction	93
4.2	Methods	96

4.2.1	Search strategy	96
4.2.2	Inclusion criteria.....	96
4.2.3	Data extraction	96
4.2.4	Analysis.....	97
4.3	Results	98
4.3.1	Overview of the field and external validity of model studies	98
4.3.2	Exploring construct validity	101
4.3.3	Exploring predictive validity	108
4.3.4	Exploring face validity.....	111
4.4	Discussion	122
4.4.1	External validity	123
4.4.2	Construct validity	125
4.4.3	Predictive validity.....	127
4.4.4	Face validity	131
4.4.5	Limitations	135
5	Text Mining as a Tool to Aid Systematic Reviews: A New Methodological Approach.....	137
5.1	Introduction.....	137
5.1.1	Motivation behind work.....	137
5.1.2	Objective of work undertaken	138
5.2	Background	138
5.2.1	Work carried out so far and possibilities for improvements	139
5.2.2	Previous work in similar fields.....	141
5.2.3	The relevance of this work to our systematic review process.....	143
5.3	Work undertaken	144
5.3.1	Approach to design.....	144
5.3.2	Design of Solution	152
5.3.3	Development of Search words.....	163

5.3.4	Performance of Solution and Evaluation	165
5.3.5	Critical analysis of results.....	169
5.3.6	Conclusions and Future Work	171
6	Methodological Quality in the Literature.....	177
6.1	Introduction	177
6.2	Methods	179
6.3	Results.....	180
6.3.1	Overall reporting of risk of bias and other quality measures by studies in the literature	180
6.3.2	Change in prevalence of reporting of risk of bias and other quality measures as a result of different factors.....	182
6.3.3	Does low quality of reporting affect outcome reported within publications?	192
6.4	Discussion.....	194
6.4.1	Change in reporting over time	196
6.4.2	Differential reporting based on study design.....	198
6.4.3	Limitations.....	199
7	Capturing all the Data of Relevance	201
7.1	Missing data due to publication and reporting biases	201
7.1.1	Methods	202
7.1.2	Results.....	203
7.2	Missing data due to limitations across time and space	205
7.2.1	Cocaine: a model of substance-induced psychosis	205
7.2.2	Are we lagging behind in light of new data: the impact of updating a search	209
7.2.3	Are we capturing all the publications of relevance in our search: the impact of alternate search terms	225
7.3	Discussion.....	241
7.3.1	Publication bias and selective reporting bias	241

7.3.2	Out-of-date search.....	243
7.3.3	Using alternative search methods.....	246
7.3.4	Limitations	249
8	Discussion	251
8.1	Models most often reported in the literature have limited relevance to psychotic disorders in the clinic	254
8.2	Assessment of animal models in the field is disordered and inconsistent	256
8.3	Lack of repetition to increase the robustness of results of individual studies 257	
8.4	Limitations in external validity of experiments	259
8.5	Limitations in face validity of developmental animal models.....	261
8.6	Limitations in construct validity of developmental animal models.....	264
8.7	Limitations in predictive validity of developmental animal models	266
8.8	Overall reporting of risk of bias and other methodological quality criteria is low 269	
8.9	Incomplete reporting of collected data	273
8.10	Lack of consistency in reporting of studies which further complicates timely review of data.....	275
8.11	Limitations	279
8.12	Conclusion.....	280
	Bibliography	285
	Appendix I. Search strategy expanded	A
	Full expansion of search terms described in Chapter 2	A
	Appendix II. Results of categorisation of the literature: Complete list of model induction methods used, behavioural outcomes measured, and treatment compounds tested	K
	Appendix III. Text mining tool: code used for program and list of expressions	M
	Code	M
	Simple word phrases.....	M

Random allocation to group.....	M
Blinded assessment of outcome.....	M
Sample size calculation.....	N
Regular expressions	O
Random allocation to group.....	O
Blinded assessment of outcome.....	O
Sample size calculation.....	O

List of Figures

Figure 1.1 Translation of research from basic science to health and practice requires us to build on building blocks of knowledge from previous domains of research	21
Figure 2.1. Venn Diagram of Disorder Classifications based on ICD-10 and DSM-5 for three major groups of disorders –psychotic, mood and anxiety disorders.	29
Figure 2.2. Screenshot of the form used in MS Access in order to screen studies identified in search against inclusion/exclusion criteria. There was a separate form for screeners 1 and 2 that looked identical.....	33
Figure 2.3. Form used for reconciliation of studies not agreed upon by screeners 1 and 2 regarding inclusion/exclusion as well as simultaneous categorization.	34
Figure 2.4. Phase II categorization form for studies that had been included by both screeners allowing the input of fields concerning reported study quality, disease model induction details, details of any treatments tested in these models and outcomes measured within the study for these models.	35
Figure 2.5 Screenshot of drop-down options for studies where more than one model was reported within the same study.....	36
Figure 2.6. Screenshot from the database used to record details about the reporting of items from our list of study quality items.....	38
Figure 2.7. Screenshot from the database used to record study identification details.	39
Figure 2.8. Screenshot from database detailing fields where data about animal model characteristics were entered.....	40
Figure 2.9. Screenshot of database showing fields used to extract data about animal husbandry and housing for specific studies.....	41
Figure 2.10. Screenshot of database where details about treatment tested and time of outcome assessment were recorded.	42
Figure 2.11. Screenshot of database showing fields where details of outcome measures would be recorded.....	43
Figure 2.12. Screenshot of database showing fields where specific outcome data measures from experiments were recorded.....	44
Figure 3.23.1Phase I categorisation of studies at abstract level.....	61
Figure 3.13.2Flow diagram of publication inclusion.....	61
Figure 3.3 Prevalence of types of model induction methods reported in the literature over time.....	65

Figure 3.4 Top 15 interventions used to induce an animal model of psychosis pharmacologically.....	66
Figure 3.5 Top 15 genetic manipulations used to induce an animal model of psychosis	66
Figure 3.6 Top 15 interventions used to induce a developmental animal model of psychosis	67
Figure 3.7 15 most widely reported lesion models in the preclinical psychosis field	68
Figure 3.8 Prevalence of human behaviours of relevance to outcomes being assessed in animal models of psychosis over time.....	70
Figure 3.9 10 most commonly reported measures of animal behaviour of relevance to psychomotor agitation	71
Figure 3.10 10 most commonly reported anxiety-like behaviours in the preclinical literature	72
Figure 3.11 10 most commonly reported behavioural measures in animal models of psychosis of relevance to sensorimotor gating	72
Figure 3.12 10 most commonly reported behavioural measurements in animal models of psychosis of relevance to deficits in learning and memory	73
Figure 3.13 15 most widely reported dopaminergic treatment drugs in the literature	74
Figure 3.14 15 most widely reported serotonergic drugs treatment drugs in the literature	75
Figure 3.15 15 most widely reported glutamatergic treatment drugs in the literature	75
Figure 4.1 Developmental methods used to induce psychosis under models of infection and adverse rearing conditions	101
Figure 4.2 Relationship between the number of administrations of prenatal infection and reported effect size in model-characterising studies	104
Figure 4.3 Treatments recorded to be tested in the literature. Number show number of experiments exploring these treatments.....	108
Figure 4.4 Prevalence of behavioural outcome measures reported model-characterising experiments of developmentally-induced animal models of schizophrenia	112
Figure 4.5 Relationship between stage of life at which behaviour is measured in animals and reported effect size in model-characterising experiments	114

Figure 4.6 Relationship between time of outcome assessment and reported effect size in model-characterising experiments	114
Figure 4.7 Prevalence of behavioural outcome measures reported model treatment-testing experiments of developmentally-induced animal models of schizophrenia	117
Figure 4.8 Prevalence of human behaviours being modelled through animal outcome measures of behaviour in model-characterising experiments	120
Figure 4.9 Prevalence of human behaviours being modelled through animal outcome measures of behaviour in treatment-testing experiments	121
Figure 5.1 Key stages of applying text mining to the process of reported study quality assessment	147
Figure 5.2 Confusion matrix showing how performance of computer tool is evaluated against human screener	150
Figure 5.3 The architecture of the program based on a REST API design that allows for complicated tasks to be separated out and performed “behind the scenes” by the back-end, while communicating only the relevant data with the front-end	153
Figure 5.4 Detailed flowchart of how the search function of the tool works – including how front-end and back-end communicate	155
Figure 5.5 Detailed flowchart of how the file check function of the tool works – note that the database update is the same as for the search function.....	156
Figure 5.6 Program start page where user is able to perform two different types of searches.....	157
Figure 5.7 Four main workflows that use four different functions initially, and then share the same update database function	158
Figure 5.8 Appearance of the user interface when results of a single search are returned and displayed	159
Figure 5.9 Appearance of the user interface when both searches are selected and executed simultaneously, and results are returned and displayed	159
Figure 5.10 User interface as it appears after the user has updated the database	160
Figure 5.11 Results displayed on page in response to the program being prompted by the user to check files	161
Figure 5.12 Screenshot of how automated assessment of measures taken to reduce the risk of bias is currently being implemented in the online platform SyRF (www.syrf.org.uk) – screenshot taken 22/08/2017	175
Figure 6.1 Prevalence of reporting of risk of bias items and other methodological quality criteria in the field over time.....	183

Figure 6.2 Prevalence of the reporting of risk of bias items and other methodological quality criteria as a percentage of total publications reported in the same year	183
Figure 6.3 Prevalence of the reporting of each item on the list by type of study....	185
Figure 6.4 Prevalence of the reporting of each item on the list by method used to induce model.....	188
Figure 6.5. Prevalence of the reporting of each item on the list by outcome measured	191
Figure 6.6 Model characterising studies reviewed here that do not report certain risk of bias items do not show statistically significantly different effects when compared to studies that do report these risk of bias items.	193
Figure 6.7 Studies that do not report random allocation to group overestimate the efficacy of treatments compared to studies that do report it. In comparison, the effect of studies reporting or not reporting blinded assessment of outcome were not found to be statistically significantly different.	194
Figure 7.1 Funnel plot of model-characterising and treatment-testing experiments	203
Figure 7.2 Egger regression plot of model-characterising and treatment-testing experiments.....	204
Figure 7.3 Effects of chronic administration of a stimulant like cocaine to animals in the laboratory (Figure adapted from Ujike, 2002).....	207
Figure 7.4 Effects of chronic abuse of cocaine in humans	207
Figure 7.5 The sensitization model of schizophrenia	208
Figure 7.6 Experimental design of acute and chronic administration paradigms of cocaine induced psychosis in animal models	212
Figure 7.7 Relationship between dose of cocaine used to induce the model and reported effect size in model-characterising studies	216
Figure 7.8 Relationship between the reporting or not reporting of randomisation and reported effect size in model-characterising studies	217
Figure 7.9 Relationship between the dose of cocaine used to induce model and reported effect size in model-characterising studies from the alternate search	219
Figure 7.10 Relationship between acute dose of cocaine used to induce the model and reported effect size in model-characterising experiments of the original search	231

Figure 7.11 Relationship between acute dose of cocaine used to induce the model and reported effect size in model-characterising studies from the alternate search.	233
Figure 7.12 Relationship between the strain of animal used and reported effect size in chronic model-characterising studies from both original and alternate searches	240
Figure 8.1 Research in basic science needs to be able to satisfy 4 key qualities of robust and reliable research in order to effectively support subsequent clinical research and practice	253
Figure 8.2 Translation of knowledge from basic science to clinical research and practice is clearly limited by four main issues, improvement of which might in turn also improve translation	281

List of Tables

Table 2.1 Variables of interest in exploring sources of heterogeneity for model-characterising and treatment exploring studies	56
Table 3.1 Prevalence of model induction method used to induce animal model of psychosis reported in the literature	63
Table 3.2 Methods used to induce model of psychosis categorised	65
Table 3.3 Human behaviours of relevance to outcomes reported to be measured in the preclinical literature	69
Table 3.4 25 most commonly reported treatments tested in animal models of psychosis in the literature	74
Table 4.1 Developmental stages of life for mice, rats and rhesus macaques	99
Table 4.2 Global estimates of effect of model on behaviour at specific stages of life	100
Table 4.3 Global estimates of effect of model on behaviour in young and adult animals	100
Table 4.4 Global estimates of efficacy for treatments in young and adult animals.	100
Table 4.5 Details of experiments describing prenatal infection models	103
Table 4.6 Global estimates of efficacy for methods used to induce the animal model	107
Table 4.7 Global estimates of effect for simple or combined model induction methods.....	107
Table 4.8 Global estimates of treatment efficacy in developmentally-induced animal models of schizophrenia	111
Table 4.9 Global estimates of effect for the three most widely reported measures of behaviour in developmentally-induced animal models of schizophrenia.....	116
Table 4.10 Global estimates of effect of behavioural outcome measures in model-characterising experiments	119
Table 4.11 Global estimates of effect when stratifying data according to human behaviours being modelled in model-characterising experiments	121
Table 4.12 Global estimates of effect when stratifying data according to human behaviours being modelled in treatment-testing experiments.....	122
Table 5.1 Confusion matrices showing number of publication when comparing predictions made by the computer and classifications made by the human reviewers	166

Table 5.2 Calculated performance measures for the search expressions developed using the dictionary-based approach	166
Table 5.3 Calculated performance measures for the search expressions developed using the rule-based approach	167
Table 5.4 Table showing level of success for each requirement that was specified at the start of the project.....	168
Table 6.1 Overall prevalence of the reporting of risk of bias items and other methodological quality criteria in the field	180
Table 7.1 Overall prevalence and potential impact of publication bias.....	203
Table 7.2 A comparison of the reporting of methodological quality items and measures taken to reduce risks of bias in studies identified in our original search, in the updated search and with all the data pooled.....	214
Table 7.3 A comparison of the data extracted from studies reporting acute cocaine-induced locomotor activity in the original search and in the updated search.....	215
Table 7.4 A comparison of the data extracted from studies characterising chronic cocaine-induced animal models in the original search and in the updated search.	221
Table 7.5 A comparison of the data extracted from studies testing treatment drugs on chronic cocaine-induced animal models in the original search and in the updated search.	224
Table 7.6 A comparison of the reporting of methodological quality items and measures taken to reduce risks of bias in studies identified in our original search, in the alternate search and with all the data pooled.....	229
Table 7.7 A comparison of the data extracted from studies reporting acute cocaine-induced locomotor activity in the original search and in the alternate search.	230
Table 7.8 A comparison of the data extracted from studies reporting chronic cocaine-induced locomotor activity in the original search and in the alternate search,	235
Table 7.9 A comparison of the data extracted from studies reporting treatment effects on chronic cocaine-induced locomotor activity in the original search and in the alternate search.....	238

Declaration

I declare that this thesis has been composed by me and the work presented here is that of my own, unless clearly stated within each section. I confirm that this work has not been submitted for any other degree or professional qualification.

Zsanett Bahor

30/10/2018

Acknowledgements

Over these past few years, I have learnt a huge amount about both neuroscience and myself, and I have become immensely passionate about the topic of mental health through my work and personal experiences. I really do hope that we will be able to better understand, recognise and help individuals with these symptoms one day. This work would not have been possible without the people who have supported me along the way.

Firstly, I would like to acknowledge my supervisor Professor Malcolm Macleod. I am indebted to you for the opportunity, for your patience and for your continued support. For knowing what to say most of the time, and for every other piece of ‘genius’ advice you have gifted me with over the years. I would also like to thank my other supervisor Dr. Emily Sena; you have not only been a mentor, but also a true friend.

Moreover, this would not have been possible without my team, CAMARADES. For all who are and have been a part of the group, those who have come and gone during my time with them- thank you for keeping me fit mentally and physically through monthly fitness challenges, gin-Fridays and cheese jokes. Kieren Egan and Hanna Vesterinen, thank you for using your expertise to do everything from helping me write my first abstract to encouraging me not to give up. Gillian Muller, from helping with analyses, to being my conference buddy; I hope that one day you figure out a way to remember all your passwords. Jing Liao, thank you for all your technological help and constant entertainment, don’t change. Sarah McCann, thank you for being my analysis troubleshooter, even outside of office hours, you are a wise, wise woman. Qianying Wang for creating the shiny app and for answering my emails about it even on weekends. Cadi Moy, thank you for all the lunchtime runs and crazy adventures, it was never the same without you. And finally, Alexandra Bannach-Brown there are too many things to mention, but I will never forget our conferences and wine clubs featuring endless scientific debates.

To everyone who I have met on this journey, at scientific conferences and other events, students in other laboratories across the world in the likes of Brazil, Denmark, the Netherlands, Germany and Australia, to mention a few – it has been my pleasure. Thank you for the amazing food, drinks and scientific discussions.

Thank you to Professor Stephen Lawrie and Professor Andrew Mackintosh for the opportunity to see the bigger picture of my work, and its importance to those

suffering from these symptoms. A special thank you to Professor Lindsay Thomson for the helpful discussions on the clinical aspects of schizophrenia. Thank you to Professor Miklós Kozlovszky for allowing me to come and learn at your laboratory in Budapest and to Viktor Kozlovszky for helping me with coding and staying sane with all the debugging. Moreover, thank you to Professor Gregers Wegener for the perspective of an animal scientist.

Thank you to my mum, who still has no idea what I am doing in my work, but who has never failed to support me despite this. Thank you to my sister who has taught me everything about motivation, perseverance and where to put and not put a comma. Thank you to Jessica Tadhunter for still proofreading my work even so many years after we graduated. And to other friends, especially Samantha Mielke, who kept me sane and encouraged me to keep going, I promise to be more social from now on.

Finally, to CDM - my coach, my personal chef and my companion for the last leg of this journey. It was a very steep mountain to climb, but it was infinitely more of an adventure with you. Thank you for believing in me on days when I didn't.

I would like to dedicate this work to Nagyus, who was not around to watch me take on this chapter of my life, but who no doubt would have been one hell of a cheerleader throughout it.

Publications and Conference Participation

Publications

Wang, Q., Liao, J., Hair, K., Bannach-Brown, A., **Bahor, Z.**, Currie, G.L., McCann, S.K., Howells, D.W., Sena, E.S., Macleod, M.R., 2018. Estimating the statistical performance of different approaches to meta-analysis of data from animal studies in identifying the impact of aspects of study design. *bioRxiv* 256776. doi:10.1101/256776

Bahor, Z., Liao, J., Macleod, M.R., Bannach-Brown, A., McCann, S.K., Wever, K.E., Thomas, J., Ottavi, T., Howells, D.W., Rice, A., Ananiadou, S., Sena, E., 2017. Risk of bias reporting in the recent animal focal cerebral ischaemia literature. *Clin. Sci.* 131, 2525–2532. doi:10.1042/CS20160722

Bahor, Z., Nunes-Fonseca, C., Thomson, L.D.G., Sena, E.S., Macleod, M.R., 2016. Improving our understanding of the in vivo modelling of psychotic disorders: A protocol for a systematic review and meta-analysis. *Evidence-based Preclin. Med.* 3, e00022. doi:10.1002/ebm2.22

Davis, C.K., Laud, P.J., **Bahor, Z.**, Rajanikant, G., Majid, A., 2016. Systematic review and stratified meta-analysis of the efficacy of carnosine in animal models of ischemic stroke. *J. Cereb. Blood Flow Metab.* 36, 1686–1694. doi:10.1177/0271678X16658302

Macleod, M.R., Lawson McLean, A., Kyriakopoulou, A., Serghiou, S., de Wilde, A., Sherratt, N., Hirst, T., Hemblade, R., **Bahor, Z.**, Nunes-Fonseca, C., Potluru, A., Thomson, A., Baginskaite, J., Egan, K., Vesterinen, H., Currie, G.L., Churilov, L., Howells, D.W., Sena, E.S., 2015a. Correction: Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLOS Biol.* 13, e1002301. doi:10.1371/journal.pbio.1002301

Macleod, M.R., Lawson McLean, A., Kyriakopoulou, A., Serghiou, S., de Wilde, A., Sherratt, N., Hirst, T., Hemblade, R., **Bahor, Z.**, Nunes-Fonseca, C., Potluru, A., Thomson, A., Baginskaite, J., Egan, K., Vesterinen, H., Currie, G.L., Churilov, L., Howells, D.W., Sena, E.S., 2015b. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLoS Biol.* 13, 1–12. doi:10.1371/journal.pbio.1002273

Presentations

Improving our Understanding of the in vivo modelling of Psychotic Disorders

Presented at:

Forensic Network Research Conference 2014 (Poster presentation),

3rd International Symposium on Systematic Reviews in Laboratory Animal Science 2014
(Platform presentation),

Brain Centre Rudolf Magnus Research Day 2014 (Poster presentation),

XIII International Conference on Translational Medicine 2015 (Platform presentation).

Data Mining as a Tool to Speed Up Systematic Reviews

Presented at:

10th FENS Forum of Neuroscience 2016 (Poster presentation),

4th International Symposium on Systematic Reviews in Laboratory Animal Science 2017
(Platform presentation),

Systematic Reviews and Meta-Analyses - Introduction and their utility in informing future research

Presented at:

Universidade Federal De Santa Catarina Pharmacology Winter course 2016 (Platform).

Can experiments of animal models of Psychosis reliably inform clinical approaches? How might we find out quicker?

Presented at:

Aarhus AUGUST Symposium in Systematic Reviews and Meta-analyses 2017 (Platform presentation)

List of Abbreviations

5-HT	5-Hydroxytryptamine (Serotonin)
95% CI	95% Confidence intervals
Ach	Acetylcholine
Ara-C	Cytosine arabinoside
CANTAB	Cambridge neuropsychological test automated battery
CNTRICS	Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia
CNV	Copy number variants
DISC1	Disrupted-in Schizophrenia
DSM-5	Diagnostic and Statistical Manual of Mental Disorders , Fifth Edition
GABA	gamma-amino butyric acid
GWAS	Genome Wide Association studies
ICD-10	International Statistical Classification of Diseases and Related Health Problems 10th Revision
IL-6	Interleukin-6
LPS	Lipopolysaccharide
LSD	Lysergic acid diethylamide
MAM	Methylazoxymethanol acetate
MATRICS	Measurement and Treatment Research to Improve Cognition in Schizophrenia
MK-801	Dizocilpine
MRI	Magnetic resonance imaging
NMD	Normalized mean difference
NMDA	N-methyl-D-aspartate
Nurr1	Nuclear receptor related-1 protein
PCP	Phencyclidine
PDF	Portable Document Format
PET	Positron emission tomography
PFC	Prefrontal cortex
Poly I:C	Polyinosinic:polycytidylic acid
PPI	Pre-pulse inhibition
SMD	Standardized mean difference
SNP	Single nucleotide polymorphism
SSC	Sample size calculation
TURNIS	Treatment Units for Research in Neurocognition in Schizophrenia
TURP	Turpentine
WHO	World Health Organization

Abstract

Psychotic disorders represent a severe category of mental disorders affecting about one percent of the population. Individuals experience a loss or distortion of contact with reality alongside other symptoms, many of which are still not adequately managed using existing treatments. While animal models of these disorders could offer insights into these disorders and potential new treatments, translation of this knowledge has so far been poor in terms of informing clinical trials and practice. The aim of this project was to improve our understanding of these pre-clinical studies and identify potential weaknesses underlying translational failure.

I carried out a systematic search of the literature to provide an unbiased summary of publications reporting animal models of schizophrenia and other psychotic disorders. From these publications, data were extracted to quantify aspects of the field including reported quality of studies, study characteristics and behavioural outcome data. The latter of these data were then used to calculate estimates of efficacy using random-effects meta-analysis.

Having identified 3847 publications of relevance, including 852 different methods used to induce the model, over 359 different outcomes tested in them and almost 946 different treatments reported to be administered. I show that a large proportion of studies use simple pharmacological interventions to induce their models of these disorders, despite the availability of models using other interventions that are arguably of higher translational relevance. I also show that the reported quality of these studies is low, and only 22% of studies report taking measures to reduce the risk of biases such as randomisation and blinding, which has been shown to affect the reliability of results drawn.

Through this work it becomes apparent that the literature is incredibly vast for studies looking at animal models of psychotic disorders and that some of the relevant work potentially overlaps with studies describing other conditions. This means that drawing reliable conclusions from these data is affected by what is made available in the literature, how it is reported and identified in a search and the time that it takes to reach these conclusions. I introduce the idea of using computer-assisted tools to overcome one of these problems in the long term.

Translation of results from studies looking at animals modelling uniquely-human psychotic disorders to clinical successes might be improved by better reporting of studies including publishing of all work carried out, labelling of studies more uniformly so that it is identifiable, better reporting of study design including improving on reporting of measures taken to reduce the risk of bias and focusing on models with greater validity to the human condition.

Lay Summary

Psychotic disorders are a group of debilitating mental disorders affecting about one percent of the population. These disorders are characterized by symptoms where individuals experience a loss or distortion of contact with reality. Despite the availability of treatments for these individuals, symptoms are not uniformly managed across different individuals and a large proportion of people remain unresponsive to these medications. Experiments using animals to model these human conditions are informative and are carried out with the aim of improving our understanding of the underlying biology of these disorders, while also providing a platform from which new potential treatments can be tested before they are taken forward to clinical trials. Unfortunately these data collected pre-clinically have not led to significant changes in clinical practice. The aim of this work was to better understand this area of research and identify the reasons underlying their weaknesses in informing clinical research.

I performed a search of the scientific literature to identify a set of publications that report experiments of animal models of all psychotic disorders including schizophrenia. Once all studies of relevance were identified, information reported within these publications was extracted. This included the reported quality of methodology, details about experimental design and finally results from these experiments. This information was then used to quantify the field and summarise some of these data to calculate overall effectiveness of a model or overall efficacy of certain treatments.

The search performed identified almost 4000 publications of relevance, including over 800 different methods used to create each animal model, over 300 different outcomes used to measure performance of these animals and almost 1000 different treatments tested in these animals. I show that about 40% of these studies use simple drug-induced methods to create their models of these disorders, despite the availability of models that are perhaps more relevant to the human condition. I also show that study quality is not well reported across studies, which has been shown previously by other research to negatively affect the reliability of results drawn. Overall, it is clear that the literature describing these experiments using animal models of psychotic disorders is very broad and that some experiments described are similar to experiments performed as part of research looking at other disorders. In order to be able to draw meaningful conclusions and use these to then inform clinical research and practise, we need to be able to identify all the relevant research that has been carried out pre-clinically, which is affected by cases where research is not published or not described in a manner expected. Moreover as there are a large number of these pre-clinical studies that have been and continue to be performed, summarising these data takes time and new tools need to be considered to be able to provide conclusions from these data faster and more up-to-date.

The potential of studies using animal models of psychotic disorders in informing clinical studies in future might be improved by better reporting of studies including publishing of all work carried out, labelling of studies more uniformly so that it is identifiable, better reporting of study quality and design and concentrating on research involving models that have more relevance to the human condition.

"If I have seen further, it is by standing upon the shoulders of giants"

Sir Isaac Newton

From a letter written to a fellow scientist, Robert Hooke (1675)

1 Introduction

1.1 The Burden of Mental Health

Mental Health is defined by the World Health Organization (WHO) as a “a state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community” (World Health Organization, 2001). While good mental health is not simply the result of the absence of a defined mental disorder, mental disorders are becoming an exponentially increasing burden on our society with about 450 million people worldwide affected (World Health Organization, 2001).

In the clinic, the diagnosis of mental health disorders is carried out using two universally established tools. These are the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) published by the WHO; and the Diagnostic and Statistical Manual of Mental Disorders, 5th Revision (DSM-5) distributed by the American Psychiatric Association (APA) (American Psychiatric Association, 2013; World Health Organization, 2016). These classification systems have some limitations due to their categorical approach to the classification of mental disorders, which fails to capture the individual differences in the severity of mental disorders (Brown and Barlow, 2005). Broadly speaking all the disorders described under mental and behavioural disorders are characterized by some mixture of atypical thoughts, emotions and behaviours and clinical profiles therefore overlap between many of these disorders so that diagnosis is often not clear-cut (Stern et al., 2008; World Health Organization, 2016). Results from scientific literature databases such as PubMed reveal that mental disorders at the cognitive–emotional interface over the last decade have received growing attention

in research. Psychotic disorders represent a severe category of these mental disorders.

1.2 Psychosis and Psychotic Disorders

Psychosis was introduced into the literature in 1841 by Canstatt and later used by Feuchtersleben from 1945 onwards to describe mental disorders that were “diseases of the personality” and therefore affected personality as a whole, eluding to “compound conditions” where the disorder was not exclusively limited to just the mind or the body (Beer, 1996; Burgy, 2008). The modern translation of the word means “any illness of the mind” (Beer, 1996) and today we recognise psychosis in itself not as a distinct medical diagnosis, but as a common feature of many mental disorders, which are characterised by a loss of contact with reality (Stern et al., 2008). These disorders represent a group of incredibly mysterious mental disorders, which are marked by disorganized thinking and perceptions that manifest themselves most illustriously through delusions and hallucinations. The overall prevalence of psychotic disorders has been reported to be about 0.4% in the United Kingdom in adults aged 16 to 64 years (Great Britain Office for National Statistics, 2002), with incidence rates being the highest for non-affective psychoses at 23 per 100 000 person-years and schizophrenia accounting for about 15 per 100 000 person-years (Kirkbride et al., 2012).

1.2.1 Nosology

Broadly speaking we can group psychotic disorders into four main diagnostic categories. These include: (1) non-affective psychotic disorders that are primarily considered to be psychotic disorders and include examples like schizophrenia and schizoaffective disorder; (2) affective psychotic disorders, or mood disorders that present with psychotic features, like bipolar disorder or major depressive disorder with psychotic features; (3) secondary psychotic disorders such as substance-induced psychotic disorders; and (4) psychotic disorders due to a general medical condition (van Os and Kapur, 2009).

Primary psychotic disorders including schizophrenia spectrum disorders are thought to occur along a spectrum, with schizoid disorder on the mild end and schizophrenia on the severe end of the spectrum (Arciniegas, 2015). Symptoms of psychosis are usually seen alongside other dimensions of neuropsychiatric

disturbances in these disorders. And therefore psychotic disorders are characterised by a diverse psychopathology and have to include abnormalities in at least one of the following domains (American Psychiatric Association, 2013):

- 1) delusions, or false beliefs that are fixed even in the context of conflicting evidence; 2) hallucinations, defined as perceptions without corresponding external stimuli;
- 3) cognitive deficits such as disorganized thinking and perception, including speech; 4) disorganized motor behaviour, including catatonia;
- and 5) negative symptoms, such as avolition and blunted emotional expression, which are more prominent in schizophrenia than other psychotic disorders.

In the clinic, these psychotic disorders are heterogeneous in their symptomology. Diagnosis is made in the clinic based on history and examination of mental state, there are no diagnostic tests (Owen et al., 2016). Schizophrenia, the most common and debilitating of the psychotic disorders, is best characterized by a triad of symptoms including positive symptoms such as hallucinations and delusions; negative symptoms such as anhedonia and avolition; and cognitive deficits, especially in memory, attention, verbal fluency and executive function (Andreasen, 1995; Savilla et al., 2008). These symptoms are required to be present for at least 6 months. If the total duration of illness is less than 6 months then it is classed as schizophreniform disorder, and if it less than a month but more than a day it is classified as a brief psychotic disorder (American Psychiatric Association, 2013). Schizoaffective disorder is characterized by a mood disturbance in the form of major depressive or manic episode occurring with other active phase symptoms of schizophrenia. Making a differential diagnosis between schizophrenia, schizoaffective disorder and depressive or bipolar disorder with psychotic features is often difficult due to the large overlap in symptomology. Delusional disorder is characterized by delusions that persist for 1 month or longer and differs from schizophrenia and schizophreniform disorder in that other characteristic symptoms of schizophrenia are absent and the disorder usually has a much later onset (American Psychiatric Association, 2013; Marek and Merchant, 2005). Psychosis can also be secondary to certain medical conditions or substances of abuse. These are categorized by delusions and hallucinations developed during or after substance intoxication or during withdrawal from substance, or where these symptoms are the

direct consequence of another medical condition (American Psychiatric Association, 2013). This syndromic approach of diagnostic criteria has been useful in categorizing disorders in the clinic, however, the assumption that these criteria map onto valid disease entities with specific causes and pathophysiology is believed to have hampered research progress over the years in these fields of research (Owen et al., 2016).

Schizophrenia, being the most heterogeneous and least understood of the psychotic disorders has seen an incredible amount of research interest over the last century. It is difficult to summarise and make sense of all this research as several hundred thousand books and papers have been published on the topic of schizophrenia. A quick search of the literature in PubMed suggests that about 6000 publications relating to schizophrenia were published in the year 2017 alone. It is also predicted that this is an underestimate as it does not account for studies not indexed or those published in foreign languages and therefore not picked up during a search of the literature. Our overall understanding of this disorder is limited by abstracts of studies that do not report clear findings, those with methodological limitations not often obvious, and findings that are rarely replicable. This makes it difficult to identify important leads within the field (Tandon et al., 2008). As the majority of pre-clinical evidence identified and reviewed in this thesis refers to experimental conditions modelling schizophrenia, I will mainly focus on what we know about schizophrenia from here on. I will discuss some highlights in our understanding based on findings in the field that are thought to potentially play a role in the development of the disorder. Unfortunately, even in the context of these findings, our understanding of the disorder still remains limited in many ways, warranting further research in the area despite the slow progress.

1.2.2 What we think we know about schizophrenia

1.2.2.1 *Clinical presentation*

In the clinic schizophrenia has a heterogeneous presentation with wide inter-individual variation in terms of disease development, clinical course of the disorder, and symptoms appearing with varying levels of prominence throughout the course of the disorder (American Psychiatric Association, 2013; Rietschel et al., 2017). The positive symptoms of the disorder relapse and remit over time; however, some individuals can have long-term psychotic symptoms. Negative symptoms and

cognitive deficits on the other hand, may persist when positive symptoms are in remission and are therefore chronic and are associated with long-term burden on social function (Owen et al., 2016). Schizophrenia primarily has a peak of onset in young adulthood, and very rarely occurs before adolescence or after middle age (American Psychiatric Association, 2013). There also seem to be sex differences whereby men have a greater lifetime risk of developing the disorder, with an earlier age of onset (i.e. by about 3-5 years), more pronounced negative symptoms and higher deficits in social functioning (Rietschel et al., 2017). Women on the other hand, are reported to experience more depressive and paranoid symptoms and usually report better outcomes (Kelly, 2006). There is also a strong link to environmental influences, whereby schizophrenia appears to be more prevalent among lower socio-economic classes and migrant and minority ethnic populations (Morgan et al., 2010), as well as being associated with urbanicity (Vassos et al., 2012).

1.2.2.2 Genetic liability and susceptibility

Most psychotic disorders have a strong genetic component as shown by a heritability rate of 81% for schizophrenia, and about 82-85% for other psychotic disorders (Cardno et al., 1999; Sullivan et al., 2003). This strong genetic component to the disorder is reiterated by the fact that biological relatives of individuals diagnosed with schizophrenia also show mild cognitive deficits (i.e. in attention, learning and memory) (Snitz et al., 2006), and are more often affected by psychiatric disorders themselves (Onstad et al., 1991).

Our understanding of the genetics underlying schizophrenia is continuously evolving. Before genome-wide association studies (GWAS), genomic research in schizophrenia focused on candidate genes (Farrell et al., 2015). However, it is now thought that most past studies of these candidate gene associations had too low statistical power to detect a true association, and thus the idea of a particular gene being associated with schizophrenia has been challenged in recent years (Farrell et al., 2015). Nevertheless, many of these early genes identified and linked to the disorder converge functionally and seem to support the hypothesis that schizophrenia develops as a result of deficits in connectivity and synaptic signalling (Harrison and Weinberger, 2005). GWAS since then have identified more than 100 distinct genome-wide significant loci linking schizophrenia to multiple common genetic variants, all with small effect sizes (Schizophrenia Working Group of the

Psychiatric Genomics Consortium et al., 2014), and they continue to find common genetic variants in individuals with schizophrenia to associate with the disorder. Many of these relate to the function and plasticity of certain neurotransmitter receptors and synapses, specifically those of N-methyl-D-aspartate receptors and glutamatergic synapses (Friston et al., 2016). Therefore, the genetic liability for schizophrenia is thought to be polygenic. Risk alleles have been identified in many different genes and many of these overlap with other psychiatric disorders (Rees et al., 2015), reiterating the limitations of stratifying psychiatric disorders according to current diagnostic criteria (Owen et al., 2016). High heritability rate suggests that the risk is mainly inherited, however, *de novo* mutations (i.e. mutations arising for the first time in a family) also occur and rates are found to be higher in individuals with schizophrenia (i.e. 5% compared to 2% in control subjects) (Malhotra et al., 2011). Next-generation DNA sequencing techniques are also uncovering single nucleotide polymorphisms (SNPs, i.e. variations in the DNA sequence by a single nucleotide) and copy number variants (CNVs, i.e. large sections of the genome are repeated or deleted) that have been implicated in the disorder (Jia et al., 2017). Ultimately, it is thought that disease susceptibility is most likely a result of the interplay between a number of different genetic changes (Friston et al., 2016). However, despite this collection of knowledge, most of these genetic discoveries have not yet been applied to clinical use of genetic diagnoses for schizophrenia (Owen et al., 2016), suggesting we still have some way to go in our understanding of the contribution of these factors.

1.2.2.3 Environmental risk factors

Other non-genetic risk factors, such as environmental influences are also thought to play a role in the aetiology of schizophrenia. For example, evidence shows that common or shared environmental influences account for about 11% of liability to schizophrenia in twins (Sullivan et al., 2003). Twin studies also show that the concordance rates for these diagnoses between identical twins is just below 50% (Cardno et al., 1999), suggesting that genetics alone don't explain the entirety of the story and do not necessarily predict the development of complex psychotic disorders like schizophrenia (Maynard et al., 2001). In fact, some research shows that environmental influences such as rearing environments can also have a strong effect on the development of symptoms of serious mental illnesses in children, contributing an additional risk to genetics (Tienari, 1991).

Environmental hits are varied and can be damaging at any time from conception to the onset of illness in early adulthood (Dean and Murray, 2005). Risk factors during early life include complications during birth, the season the birth occurs in, infections to the mother affecting the foetus *in utero* or the baby postnatally, malnutrition of the mother, stress to the mother and any other insults potentially affecting brain development (Dean and Murray, 2005). These observations are the main basis for the neurodevelopmental hypothesis of schizophrenia (Murray and Lewis, 1987), where abnormalities in neural development in utero starting as early as late first trimester, can lead to the activation of pathological changes in later life (Fatemi and Folsom, 2009). In childhood, factors such as early experiences of adverse upbringing, abuse or even head injury can affect the development of the brain and increase the risk of schizophrenia seen in later life. Triggers in later life might include drug abuse, chronic stress and other traumatic life events (Dean and Murray, 2005).

While these factors are correlated with the development of the disorder, they don't sufficiently predict the onset of the disorder and therefore it is thought that both genetics and environmental factors play a role in the development of the disorder (Maynard et al., 2001). Initially described as the "two-hit hypothesis" of the disorders, this theory implied that early disruptions to the central nervous system, such as an individual's genetic makeup establishes a sort of predisposition to the disorder, making them more vulnerable to another insult from the environment leading to the development of symptoms (Maynard et al., 2001). More modern versions of this hypothesis recognise that genetic, environmental and other vulnerability factors are likely to be cumulative and interactive with each other in a more complex way than previously explained by the original binary two-hit model. It is thought that the development of schizophrenia is likely a result of a multi-hit threshold model where key neurodevelopmental milestones are affected (Davis et al., 2016).

Our understanding of how these factors interact exactly are still limited. While there have been several theories on the possible mechanisms underlying the development and manifestation of the disorder, no single mechanism has been established that directly correlates with the onset of the disorder.

1.2.2.4 Structural and Functional abnormalities of the schizophrenic brain

The nature of pathological abnormalities in the brain and their exact pathogenic mechanisms underlying schizophrenia are not yet completely understood, however, there have been a number of robust observations made based on post-mortem tissue studies, as well as neuroimaging and molecular biology studies. Overall, it is thought that a set of multiple, subtle changes arise affecting the neural circuitry in the brain (Harrison and Weinberger, 2005).

Research shows that cerebral volume is reduced and total ventricular volume is greater in chronic schizophrenic individuals (Wright et al., 2000). For example, about 80% of magnetic resonance imaging (MRI) studies show an enlargement of the lateral ventricles, and about 73% of studies show third ventricular enlargement in individuals diagnosed with schizophrenia. This enlargement of lateral ventricles is thought to imply tissue loss in surrounding brain regions or failures in normal development (Shenton et al., 2001).

In addition, there seems to be a loss of cortical grey matter without cell death (Sekar et al., 2016), which has been found to be localized to mainly frontal and temporal areas of the brain (Kuperberg et al., 2003). These regions are involved in episodic memory, auditory information processing, short term memory acquisition and decision making (Karlsgodt et al., 2010). This abnormality is likely a result of abnormal thinning in cortical volume (Cannon et al., 2015) and reduced dendritic spine density in pyramidal neurons of the cortex (Garey et al., 1998). These are often targets of afferents from subcortical structures such as the thalamus, which has also been reported to be reduced in volume in schizophrenic individuals (Brickman et al., 2004). The thalamus connects multiple regions in the cortex and is thought to play a role in gating of sensory input to the cortex, which is negatively affected in schizophrenia (Rao et al., 2010). Ultimately, reductions in both grey matter volume and thalamus volume have been linked to more pronounced negative symptoms in schizophrenic individuals (Anderson et al., 2002; Rao et al., 2010).

In addition to grey matter changes, there are also disruptions in white matter integrity within the brains of individuals with schizophrenia. White matter volume has been shown to be reduced in both first-episode and chronic schizophrenic individuals, and these abnormalities seem to be localized primarily to white matter in frontal and temporal regions (Kubicki et al., 2005). Greater disruptions in white

matter are associated with a greater severity in symptoms across the schizophrenia spectrum (Lener et al., 2015)

Medial temporal lobe structures including the amygdala-hippocampal complex and parahippocampal gyrus are also reduced in volume, as observed in both first-episode schizophrenia and chronic schizophrenia (Shenton et al., 2001). These structures are involved in associative and retrieval processes in memory, which are thought to underlie aberrant auditory and language processing functions seen in schizophrenia. Reductions in the volume of the hippocampus seem to correlate with a decline in reduced verbal and spatial memory performance (Allen et al., 2016), and an earlier onset of psychosis (Stefanis et al., 1999). Anatomically, hippocampal neurons in the pyramidal cell layer are normal in schizophrenia, however studies show that there is a reduction in the number of parvalbumin-positive interneurons in the CA1 and CA4 sectors and somatostatin-positive interneurons in all three hippocampal sectors. These changes are thought to be linked to psychotic symptoms and deficits in memory and other hippocampal functions (Konradi et al., 2011). Changes in hippocampal volume seem especially influenced by environmental factors (Stefanis et al., 1999).

Activation patterns in certain brain regions are also different in individuals with schizophrenia. For example, the dorsolateral prefrontal cortex has been shown to be activated to a lesser degree in schizophrenic individuals compared to healthy individuals during the performance of working memory and executive tasks (Glahn et al., 2005). This supports the theory of hypofrontality, or the reduced physiological activity in the prefrontal cortex (PFC) of the brain, seen in brain-imaging studies of the frontal and temporal lobes in schizophrenia (Davidson and Heinrichs, 2003). In comparison to this, imaging studies show an increase in activation of the anterior cingulate cortex and left frontal pole regions of the cortex (Glahn et al., 2005). This complex pattern of hypo- and hyper-activation of different brain areas supports the proposal that normal functional connectivity in the frontal and limbic structures is disturbed in schizophrenia, making it a syndrome of dysconnectivity (Glahn et al., 2005).

Despite robust evidence establishing associated structural abnormalities in the brains of schizophrenic individuals, it is sometimes unclear at what stage of disease development they occur and how they develop over time (Dietsche et al., 2017). Some evidence shows that many structural changes such as lateral and third

ventricles enlargement, overall brain and hippocampal size reduction and white matter abnormalities are already seen at early stages of or stages leading up to the disorder (Dean et al., 2016; Vita et al., 2006). This implies that these features are not secondary to disease progression or seen as a result of anti-psychotic medication and that an aberrant course of brain development likely contributes to the onset of the disease (Karlsgodt et al., 2010). In comparison, however, there is no significant evidence to show that some other structural changes associated with schizophrenia such as a reduction in temporal lobe and amygdala volumes are present at these early stages of disease progression (Vita et al., 2006). Evidence shows that there is a significant reduction in the volume of the brain in schizophrenic individuals which occurs both before and after maximum brain volume attainment (Woods et al., 2005), implying that both aberrations in early development and at later developmental stages can negatively affect the brain.

Despite these findings, no single functional or anatomical abnormalities have been identified that are specific to schizophrenia separating it from other disorders (Owen et al., 2016), limiting the applicability of these findings in the clinic.

1.2.2.5 Neurochemical mechanisms

Several influential hypotheses of schizophrenia exist regarding the neurobiology underpinning schizophrenia, implicating different neurotransmitter systems. The two most influential hypotheses involve a dysfunction of the dopaminergic and glutamatergic systems; however, serotonin and other neurotransmitters are also thought to be potentially involved.

1.2.2.5.1 Dopamine

The dopamine hypothesis initially came from observations that dopamine agonists like amphetamine which increase extracellular dopamine concentrations exacerbated psychotic symptoms in individuals with schizophrenia (Lieberman et al., 1987) and were able to induce acute psychotic symptoms in healthy individuals (Bell, 1973). Moreover, most anti-psychotic medication block D2 dopamine receptors and their clinical efficacy in controlling hallucinations and delusions correlates with their affinity for the receptor (Seeman et al., 1976). Finally, drugs that deplete dopamine levels such as reserpine reduce psychotic symptoms (Howes and Kapur, 2009). Molecular imaging studies using radiotracers have found dopamine

release in schizophrenic individuals to be higher than in control subjects, which seems to directly relate to the degree of psychotic symptom severity seen in these individuals (Pogarell et al., 2012). More specifically, it is thought that increased release of dopamine in subcortical structures leads to increased dopamine D2 receptor stimulation (Abi-Dargham et al., 2000), and in turn positive symptoms, which are thought to be mediated through a disturbed cortical pathway through the nucleus accumbens (Brisch et al., 2014). Not surprisingly, these symptoms respond better to current anti-psychotic medication, all of which as mentioned, show antagonistic activity at dopamine D2 receptors (Abi-Dargham et al., 2000). This dopamine circuit has been implicated in emotional and motivational processing and plays a role in encoding and expressing salient learning and memory formation (Laviolette, 2007).

The negative symptoms of schizophrenia on the other hand, have been attributed to a reduction in dopamine D1 receptor activation in the PFC and nucleus (Abi-Dargham et al., 2000; Brisch et al., 2014). The PFC is involved in executive functioning, which has been shown to be impaired in individuals diagnosed with schizophrenia (Weinberger and Gallhofer, 1997) and patients with dorsomedial PFC lesions show behaviours similar to negative symptoms seen in schizophrenic individuals (Orellana and Slachevsky, 2013). Dopamine D3 receptors are also thought to possibly be involved in these latter group of symptoms as overexpression of them in the striatum of mice disrupts motivation (Simpson et al., 2014).

As a whole, the dopaminergic system is very vulnerable to outside factors such as disturbances during and around birth and development (Laviolette, 2007). Evidence shows that certain environmental factors can dysregulate the dopamine system to cause psychotic symptoms especially in individuals who already vulnerable (Stokes et al., 2013).

Nevertheless, there are some limitations to the dopamine hypothesis. For example, psychosis-like symptoms can also be induced in healthy individuals and exacerbated in schizophrenic individuals by drugs that do not directly interact with dopaminergic receptors. These include glutamate antagonists, anti-cholinergics and agonists of the cannabinoid CB1 receptor system (Laviolette, 2007). Similarly, many individuals experiencing schizophrenic symptoms do not respond well to mainly dopaminergic antipsychotic drugs or any other manipulations that involve the depletion of presynaptic dopamine (Remington et al., 2012). Finally, drugs that

modulate dopaminergic pathways and affect the presentation of psychotic symptoms such as reserpine and amphetamine, also affect other monoamines in the brain (Howes et al., 2015). Therefore, dopaminergic dysfunction is not thought to be able to fully explain the abnormal neurobiology of schizophrenia.

1.2.2.5.2 Glutamate and GABA

Some research suggests that the dopamine system is in fact “normal” in its configuration, but something else is abnormally driving the dysregulation of it in the brains of individuals with schizophrenia (Grace, 2012). The glutamatergic system is another system that has attracted a lot of attention in recent years. Glutamate signalling is involved in synaptic plasticity and cortical processing (Steeds et al., 2015), through the ability of glutamate receptors to stimulate neurite outgrowth, synaptogenesis and maturation of the synapse during development of the brain (Gaur et al., 2008). Evidence for the involvement of glutamate in schizophrenia pathophysiology comes from observations that drugs that manipulate glutamate transmission, especially *N*-methyl-d-aspartic acid (NMDA) receptor-interacting ones, like dissociative anaesthetics that block the receptor, are able to induce psychotomimetic states including negative symptoms and cognitive impairments that closely resemble schizophrenia (Javitt and Zukin, 1991). Moreover, a reduction in NMDA receptor binding has been shown in medication-free schizophrenia patients (Pilowsky et al., 2006), and enhancing NMDA receptor function has been shown to reduce negative symptoms and positively affect cognitive deficits in schizophrenic subjects (Goff and Coyle, 2001). This has led to the NMDA receptor hypofunction hypothesis of schizophrenia.

In addition to glutamate, another key player of this hypothesis is the inhibitory neurotransmitter, gamma aminobutyric acid (GABA). Decreased GABAergic signalling in the PFC has been reported consistently in post-mortem schizophrenia compared to healthy individuals (Lodge et al., 2009). This seems to be restricted mainly to the calcium-binding protein parvalbumin-containing class of GABAergic interneurons (Lewis et al., 2005) and this dysfunction in the dorsolateral PFC has been linked to deficits in working memory (Lewis et al., 2004).

It is thought that an excessive glutamate output in the glutamatergic pathway projecting from cortical pyramidal neurons through GABAergic inhibitory interneurons to dopaminergic neurons in the ventral tegmental area (VTA) leads to

overactivation of this communication pathway contributing to a hyperdopaminergic state in the mesolimbic area. This is thought to underlie the positive symptoms of schizophrenia (Homayoun and Moghaddam, 2007; Marsman et al., 2014; Stahl and Muntner, 2013). In support of this theory is the observation that administration of ketamine, an NMDA receptor antagonist, in healthy individuals leads to dopaminergic dysregulation, similar to that seen in schizophrenia (Kegeles et al., 2000).

Additionally, the mesolimbic area also gets input from the hippocampus, which can also similarly cause a hyperdopaminergic state in the mesolimbic area (Stahl and Muntner, 2013). In support of this, research has found evidence for a decreased density of parvalbumin-containing interneurons in the hippocampus (Zhang and Reynolds, 2002), as well as increased hippocampal activity, which in turn seems to correlate with psychotic symptoms, especially auditory hallucinations (Heckers, 2001; Silbersweig et al., 1995). It has been proposed that hippocampal hyperactivity could drive the dopaminergic system to be hyper-responsive to stimuli and enter a state called aberrant salience (Kapur and Mamo, 2003), where all stimuli are treated similarly and given maximal attention and reaction (Grace, 2012). This impairment in information processing means that individuals affected have an inability to ignore irrelevant stimuli as often seen in schizophrenic individuals (Heckers, 2001). Evidence shows that stress, drug abuse and other environmental factors can also affect the hippocampus (Battistella et al., 2014; Grace, 2012; Jacobus and Tapert, 2014; Mondelli et al., 2010).

Negative symptoms and cognitive deficits characteristic of schizophrenia, can also be explained to a certain degree by glutamate. It is proposed that an overactivation of a different population of glutamate neurons negatively modulates dopamine neurons in the VTA which project to the PFC, leading to the hypoactive state in the mesocortical dopamine pathway (Stahl and Muntner, 2013).

1.2.2.5.3 Serotonin

Serotonin (or 5-hydroxytryptamine; 5-HT) has also been suggested to potentially play a role in the formation of psychotic symptoms (Steeds et al., 2015). 5-HT_{2A} receptor agonists, like the hallucinogens D-lysergic acid diethylamide (LSD) and psilocybin, induce symptoms similar to those seen in first episode schizophrenic individuals, which include agitation, anxiety and hallucinations (Steeds et al., 2015).

These drugs are also able to disrupt pre-pulse inhibition (PPI) of the startle reflex response, which is a phenomenon observed in schizophrenic individuals and serves as an operational measure of sensorimotor gating (Halberstadt and Geyer, 2010). Moreover, many atypical anti-psychotics such as clozapine, risperidone and olanzapine also have additional modulating activity on the serotonergic system by antagonising 5-HT_{2A} receptors (Ichikawa et al., 2001).

Serotonergic hallucinogens enhance glutamatergic transmission, through the activation of 5-HT_{2A} (i.e. subtype of the 5-HT₂ serotonin receptor) stimulation in the PFC (Aghajanian and Marek, 2000). This and the fact that blockade of 5-HT_{2A} receptors reverses the effects of NMDA receptor antagonists, suggests the involvement of serotonergic mechanisms in the NMDA receptor hypofunction model of psychosis (Steeds et al., 2015). It is thought that atypical anti-psychotics with 5-HT_{2A} and D₂ receptor antagonistic activities lead to distal activation of the mesocortical pathway and thus increased dopamine release in the medial PFC, through the promotion of 5-HT_{1A} stimulation (Celada et al., 2013). As a hypoactive dopaminergic system in the PFC has been implicated in the formation of negative symptoms and cognitive deficits, 5-HT_{2A} receptor antagonists have been suggested to have therapeutic value for negative symptoms and cognitive deficits in individuals with schizophrenia (Ichikawa et al., 2001; Roth et al., 2004). Clozapine, an atypical anti-psychotic often suggested to be superior to other anti-psychotics and recommended for treatment-resistant schizophrenic individuals, displays agonistic properties at the 5-HT_{1A} receptor and has in some cases been shown to have an effect on both positive and negative symptoms of schizophrenia (Celada et al., 2013).

1.2.2.5.4 Cholinergic system

It is also hypothesized that misregulated, cortical cholinergic inputs play a role in the development and manifestation of attentional deficits seen during positive symptoms of schizophrenia. Its dysregulation is a correlate of the abnormal mesolimbic dopaminergic transmission underlying positive symptoms (Moore et al., 1999; Sarter et al., 2005). Dopamine agonists have been shown to increase acetylcholine outflow in the cortex (Day and Fibiger, 1993). Normally acetylcholine (ACh) release is vital in the PFC for normal cognitive functioning, and the cholinergic system plays a role in information processing and cognition (Steeds et al., 2015). Changes in markers for the ACh and levels of the ACh synthesizing enzyme, choline acetyltransferase

(ChAT) have been frequently reported in individuals with schizophrenia (Terry and Jr, 2008). Muscarinic M1 and M2 receptors that play a role in neuronal excitability, synaptic plasticity and learning and memory, seem to be especially affected. The literature also shows that administration of muscarinic antagonists like scopolamine and atropine can induce a psychotic state which is similar to the symptoms seen in schizophrenia (Steeds et al., 2015; Terry and Jr, 2008). In addition to this a reduction in the $\alpha 7$ subtype of nicotinic Ach receptors is also observed in many cases of schizophrenia and has been linked to sensory gating deficits (Freedman et al., 2005). Interestingly, abuse of anticholinergic drugs is widespread among schizophrenic individuals who report functional benefits, including improvements in negative symptoms and cognitive abilities as a result of taking these drugs (Koukoulis et al., 2017; Wells et al., 1989).

Ultimately, schizophrenia is thought to involve a disruption of neuronal connectivity in turn causing a deficit in the coordinated processing of information (Gaspar et al., 2009). Our current understanding is that a set of processes at a functional level, including a number of subtly altered networks throughout the brain contribute to reductions in synaptic activity and integration, and thus a disruption in functional dysconnectivity and miscommunication between different brain areas (Cannon et al., 2015; Pettersson-Yeo et al., 2011). A prevailing view is that *de novo* or inherited mutations in one or multiple genes cause a failure of neuromodulation, which arises as a result of aberrant interactions of neuromodulator systems involving neurotransmitters discussed above. Because of how closely neurotransmitter systems in the brain are connected with one another, it is difficult to tease out which changes seen in neurobiology are primary, secondary or even tertiary signs of the disorder and therefore the exact mechanisms behind schizophrenia remain unclear (Zink and Correll, 2015). With genes of low penetrance, environmental factors are likely to also play a role in the development of the disorder, whereas much rarer high penetrance genes might be able to cause manifestation of the disease early on leading to a more severe form of the illness (Ahn et al., 2014; Friston et al., 2016). This molecular pathology leads to an inability to process sensory evidence appropriately and can lead to false inferences in the form of hallucinations and delusions (Friston et al., 2016). Anatomical changes described above are thought to develop overtime (van Haren et al., 2011).

1.2.3 Clinical management

1.2.3.1 Current treatment of psychotic disorders and weaknesses

Pharmacological treatment is recognised as the foundation of the clinical management of individuals who experience psychotic symptoms; however, psychotherapies like cognitive behavioural therapy (CBT) have also been shown to be effective in some cases, especially when used in combination with pharmacological medication (Miyamoto et al., 2005). The serendipitous discovery of the psychotropic drug chlorpromazine in the 1950s paved the way for pharmacological treatment of psychosis and led to development of first generation or “typical” anti-psychotics; and later, in the 1990s, of second generation or “atypical” anti-psychotics (Geyer et al., 2012). Both groups of drugs bring a burden of serious side effects and those taking these medications are observed to have extrapyramidal symptoms with first-generation antipsychotics and weight gain and metabolic syndrome with second-generation antipsychotics (Leucht et al., 2009). Clinical observations suggest that in general second-generation antipsychotics have little additional value compared with first-generation antipsychotics (Gupta, 2010; Leucht et al., 2009). Current NICE guidelines suggest that selection of treatment should be based on assessment of the balance between benefits and adverse effects for each individual (National Collaborating Centre for Mental Health (Great Britain), 2014). Unfortunately, over 60 years on from the identification of the first anti-psychotic, chlorpromazine, successive pharmacotherapies since chlorpromazine have largely been modifications of early drugs and current antipsychotic treatments still converge on the same mechanism of dopamine D2 blockade for the management of psychosis (Kapur and Mamo, 2003; Zink and Correll, 2015). Our limited understanding of the pathophysiology of these disorders has hindered the development of novel treatments for psychotic disorders with no real, major breakthroughs having been made in recent decades.

Despite available treatment options, the management of psychotic disorders continues to be problematic, stemming from a number of seemingly unmet therapeutic needs. National statistics report that just two-thirds of people identified as having had a psychotic episode in 2007 in the UK received medication, counselling or any other form of treatment (Great Britain Department of Health, 2007). Moreover, one of the biggest issues with current treatment options are that there are a great proportion of individuals experiencing psychosis who are

unresponsive to medication available (Kapur and Mamo, 2003). In fact, the James Lind Alliance has established treatment of those experiencing schizophrenia that are unresponsive to treatment as the number one research priority in terms of treatment uncertainty (Lloyd and White, 2011). Recent estimations report that about 30% of patients respond to anti-psychotic treatment well and meet full remission criteria, while another 30% only show partial response to the drugs available and about 20-30% do not respond to medication at all (Steeds et al., 2015). It is widely accepted that current available medication only manages symptoms, and of these mainly positive symptoms, however, there is no evidence to show that these medications also correct the underlying biological problem (Kahn et al., 2015). In fact another central criticism of current medication is that there is little proof indicating that these anti-psychotics alleviate much other than the psychosis, which is especially problematic for individuals who also experience negative symptoms and cognitive deficits alongside the positive symptoms of psychosis, commonly seen in schizophrenia (Geyer et al., 2012). Moreover while initially it was thought second generation anti-psychotics could replace the burdensome extra-pyramidal side effects of first generation anti-psychotics, it soon became apparent that previous side effects had just been replaced with newer ones (Leucht et al., 2009). In fact, evidence from the UK Database of Uncertainties about the Effects of Treatments show that many patients worry about the side effects of current treatments especially in the long term (<https://www.library.nhs.uk/duets/>). Perhaps unsurprisingly, possibly partly due to some of these factors, it is estimated that around half of schizophrenia patients don't adhere to treatment recommendations (Gibson et al., 2013). This data suggests that new treatments should be tested in consideration of what outcomes matter the most to the people that will ultimately end up receiving these treatments (Lloyd and White, 2011).

1.2.4 Challenges in developing new treatments for psychotic disorders

The odyssey of drug development typically involves a slow process of basic research and discovery, testing and development of promising therapeutic compounds in animal models of the human deficit, and assessment of the compound in three stringent phases of clinical trials. This is followed by an anticipative approval by the appropriate agency (i.e. US Food and Drug Administration (FDA) or regional regulatory bodies in the European Union) (DiMasi et al., 2003; Steinmetz and Spack, 2009). This is usually an incredibly lengthy and costly process that usually takes years, and costs billions of dollars (Morgan et al.,

2011; Van Norman, 2016). A recent study estimated that the likelihood of approval by the FDA for a new drug entering its first phase of clinical trials is on average about 10.4% (Hay et al., 2014), or about one in nine interventions that make it through the development process and go on to be approved by European or US regulatory bodies (Kola and Landis, 2004). Most compounds fail at Phase II and Phase III of development, however, it has also been shown that about 30% of drugs designed for treatment of the central nervous system (CNS) fail at registration (Kola and Landis, 2004).

As mentioned, there has been very little progress in terms of new drug development for schizophrenia in recent years and it is clear that this is somewhat as a result of our still limited understanding of the underlying mechanisms of schizophrenia. One major limitation of progress in the field is that many findings associated with the disorder are not specific to schizophrenia and any functional and chemical changes are also associated with other psychiatric disorders, making it very difficult to thus use these observations as predictors of the disorder (Maynard et al., 2001).

Moreover, there is a large extent of variability and heterogeneity between the clinical profiles of different individuals diagnosed with schizophrenia. These issues make it difficult to predict outcome in any given individual and develop treatments effective for all (Harrison and Weinberger, 2005). Currently no biomarker exists for schizophrenia, but it is also highly unlikely in the context of this clinical heterogeneity of the disorder that a single biomarker will be able to account for the multiple underlying pathophysiological processes that are currently thought to play a role in the development of schizophrenia (Weickert et al., 2013). Individuals who experience psychotic disorders such as schizophrenia have been shown to have an 8.5-fold higher risk of suicide than when compared to the general population (Harris and Barraclough, 1998). Moreover, symptoms of schizophrenia are usually associated with a general low quality of life which is not often significantly improved by newer anti-psychotic treatments (Bobes et al., 2007), therefore current management of these disorders is clearly suboptimal and this needs to change.

1.2.5 Pre-clinical studies and the role of animal models

Naturally one of the biggest concerns when releasing a new drug into the marketplace is that they are safe and effective for the purpose that they were intended for. Low attrition rates in clinical drug development can be mainly attributed to inadequate efficacy and clinical safety (Kola and Landis, 2004; Arrowsmith,

2011). These are both factors that rely on pre-clinical studies involving animal models, therefore suggesting that current animal model experiments lack predictive value (Kola and Landis, 2004)

An animal model in science is considered to be an experimental preparation that has been developed for the purpose of studying a condition in the same or a different species (Geyer and Markou, 2000). Animal models are invaluable tools in neuropsychiatric research as they provide an environment where disease progression can be monitored at a much higher rate as opposed to humans. This allows researchers to evaluate underlying structural and neurochemical changes that lead to the brain abnormality causing the disorder (Jones et al., 2011). For the purposes of novel drug development, pre-clinical studies allow for the investigation of a proposed therapeutic compound's safety, efficacy and its toxicity in animal models, before they are given to humans. These experiments can also aid the design of the first phase of clinical trials in humans, by providing information about dosage to be evaluated and for example what signs to look for in humans to evaluate safety of the drug (Lo et al., 2009). It is imperative to keep in mind that extrapolating evidence from animals to humans is rarely straightforward and there are a number of limitations, including poor design, conduct of experiments as well as the obvious issue of lack of generalisability between the two species (Bracken, 2009; van der Worp et al., 2010)

Two ways in which researchers have attempted to create models of validity for psychiatric disorders is by reiterating the behavioural and cognitive abnormalities that are seen in the clinical phenotype of the disorder or by replicating the relevant neural, neurochemical, molecular or anatomical aspects of the disorder in question (Fernando and Robbins, 2011). In neuropsychiatric research we can broadly cluster models into four different groups: pharmacological-, genetic-, developmental- and lesion-induced models (Jones et al., 2011). Modelling disorders in animals is especially difficult for psychotic disorders, as these are heterogeneous in their symptomology and often present with high levels of co-morbidity (Fernando and Robbins, 2011; Geyer and Moghaddam, 2002). For this reason, more often than not, emphasis tends to be put on modelling symptoms rather than a disorder *per se* (Fernando and Robbins, 2011). While it is not clear that a rodent can experience psychosis as we understand it in humans (Powell and Miyakawa, 2006), such models may provide useful tools to study these complex human disorders and their

underlying structural and neurochemical changes, while also providing an assay to screen novel therapeutics in (Jones et al., 2011).

Animal models are thought to have translational relevance to the symptoms of the human disorder in question, if they have construct, face and predictive validity for the disorder being modelled (Jones et al., 2011). These criteria suggest that animal models should have etiologic validity (i.e. etiologic relevance of the methods by which a model is created), ethological validity (i.e. observable outcomes of relevance), and pharmacological validity (i.e. response to treatment predictive of effects in humans) for the disorder, respectively (Belzung and Lemoine, 2011; Nestler and Hyman, 2010). Of course, the validity of an animal model is often subjective and it is suggested that greater agreement on how to judge validators would improve research in the field (Nestler and Hyman, 2010). There is also some ambiguity to the definition of some of these validities and how they are used to judge animal models. Throughout this thesis, I will define etiologic, or construct validity, as the theoretical relevance and rationale of the methods used to construct the model to the disease; face or ethological validity as the measure of how well the model constructed manages to recapitulate the pathophysiology of the human disease; and predictive validity as the ability for an animal model to respond to treatments in the same way as humans would. By these definitions an animal model with high construct validity to a disorder would be created through the same etiologic processes as those that underlie the disorder, one with strong face validity would display anatomical, biochemical, neuropathological and behavioural features similar to those in humans, and finally one with good predictive validity would be able to predict the effect of a treatment in humans with the disorder (Belzung and Lemoine, 2011; Nestler and Hyman, 2010; Varga et al., 2010). Overall, despite many great appraisals of pre-clinical models published over the last two decades in the literature, there has not been any sort of quantification of the animal research field.

1.2.6 Translation of knowledge from basic science to the clinic and its failure

Translational medicine is the process of using information obtained in one research domain to inform and guide research in a different research domain (Figure 1.1). This term has gained popularity in the last decade as we are now beginning to question more and more why the large expansion of basic biomedical research has not seemed to make a considerable benefit to medical practice. While animal studies are invaluable in terms of informing subsequent research domains and

eventually healthcare practice, evidence shows that only about a third of highly cited animal studies (i.e. with more than 500 citations) are subsequently carried forward and translated to human randomized trials, and an estimated 11% of these interventions replicated in humans go on to be approved for the clinic (Hackam and Redelmeier, 2006). Across a number of different research fields, there seems to be an obvious breakdown in the drug development process between research done in animal models and successful drug candidates that are carried forward into clinical trials to be approved for the treatment of human patients. Many candidate treatments developed for schizophrenia have also turned out to be false positives (i.e. they have shown efficacy in animal models, but not in clinical trials) (Moore, 2010).

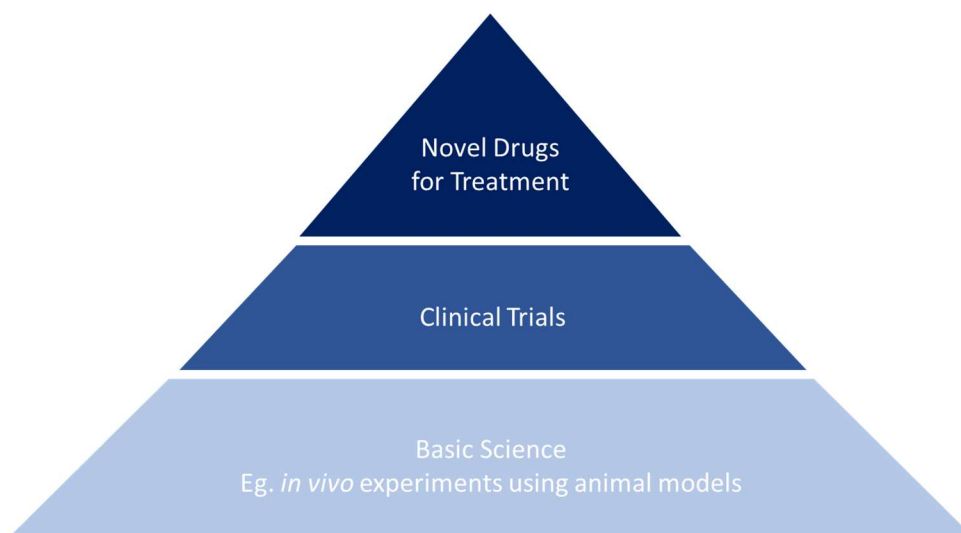


Figure 1.1 Translation of research from basic science to health and practice requires us to build on building blocks of knowledge

This raises the fundamental question: what can we blame for this apparent attrition rate between bench and bedside in CNS disorders research? One perspective is that results from both domains of research are accurate, but human physiology and disease are not sufficiently exemplified by animal models and therefore animal models fail to replicate human disease with sufficient fidelity to predict efficacy of treatments in humans (van der Staay et al., 2009). Alternatively, the animal literature is affected by biases in the design, conduct and reporting of experiments so that they provide an incomplete picture of the overall physiology (Ioannidis, 2012; van der Worp et al., 2010). Previous analyses of pre-clinical studies from a number of neurological diseases suggest that the value of research using animal models is

marred by issues with inadequate methodological quality of studies, limited construct validity of models used, and by limited external validity of experimental design applied (van der Worp et al., 2010). Overall, research is considered to be informative and of high-quality, when experiments are well-designed in advance, when they are rigorously carried out during, and results obtained are analysed correctly afterwards (Samuel et al., 2016). How valid a study is, is determined firstly, by whether the question a study is asking is answered 'correctly' and how confident we can be in the data presented. This eludes to the quality of the methodology and whether it is in general free from bias (i.e. termed "internal validity") (Krauth et al., 2013). Secondly, we need to consider whether the research question being asked is appropriate to begin with. This is termed as "external validity" and looks at whether the study is fit for the purpose it has been designed (Higgins and Green, 2008), and whether we are able to make reasonable extrapolations to humans (Geyer and Markou, 2000). Even if studies are without the possibility of bias, translation of results from animal studies may still fail if there are differences between the animal model paradigm and either the clinical picture of the disorder or clinical trials testing the proposed treatment (van der Worp et al., 2010).

These are both importantly recognized dimensions of validity for clinical studies and therefore are also of relevance to *in vivo* research. Weaknesses in these measures could lead to translational failure between research at the bench and bedside as they can lead to false interpretations of the data (Higgins and Green, 2008). Moreover, for research to be informative, the "completeness of reporting" (Moher et al., 2015) should also be of a high standard. This is essential in science if we are to build on previous observations and drive scientific progress (Landis et al., 2012).

1.2.7 Systematic reviews and their utility

To elucidate what might be to blame for this translational failure between pre-clinical and clinical research, it is clear we need a more systematic approach to evaluating animal research. A systematic review is a review that allows for the identification, synthesis and analysis of all available research in the literature in relation to a given research question (de Vries et al., 2014). It allows for the synthesis of the state of current knowledge so that it can guide future research in a more objective way than narrative reviews could, by putting any future research in the context of existing knowledge (Higgins and Green, 2008).

Single studies can rarely offer the opportunity to see the entire biological picture. By combining information from individual studies through a meta-analysis, if appropriate, power and precision of results can be increased (Noordzij et al., 2009). Systematic reviews are thought to be transparent and comprehensive and can generally offer a less biased picture than narrative reviews, for example, as they can minimise subjective selection bias of evidence. As a result, they have the potential to maximise the research that has been so far carried out in order to reduce the number of resources wasted, and replace and refine experiments so that they have higher informative value. This in turn better aids the bridging of knowledge from basic to clinical sciences.

While systematic reviews are promising and clearly useful, at current the process of distillation of knowledge using this tool is confounded by a number of both uncontrollable and controllable factors. Essentially a review can only take into account the literature that is available and reported and the part of the literature that it manages to capture. This includes important and often uncontrollable confounding factors in the review process such as poor or incomplete reporting of experimental designs, and exclusion of relevant data from a review due to data from a research study remaining unpublished or indexed in a way, which makes it unidentifiable using a certain set of search terms. Moreover it is possible to introduce systematic flaws and potential biases into the systematic review process through limitations in design and conduct of these reviews. Novel, emerging tools such as the ROBIS tool will allow us to assess these risks of bias in systematic reviews more efficiently in future (Whiting et al., 2016). Ultimately, all of these limitations can lead to a distortion of the true biological picture when summarising evidence, even when summarised in a systematic way (Rosenthal, 1979; Sena et al., 2010).

Nevertheless, by reviewing the literature on *in vivo* experiments, we can assess and critically appraise studies in the literature, to shed light on potential weaknesses in the pre-clinical field and where design and conduct of future experiments can be improved upon (de Vries et al., 2014). This is only possible if we have a bird's eye view of what is out there already. This overview then has the potential to inform what we think we know; both about how these disorders are modelled in animals, and about how they support clinical drug development.

1.3 Aim and Objectives

In order to begin to explore and unravel potential reasons for the limited progress in translation of research from animal studies of psychotic disorders to clinical research and practise, I wanted to initially improve our understanding of the preclinical research field and the role that animal models play in the drug discovery process. The work presented here was carried out with the aim of providing a transparent and comprehensive summary of the use of animal models of psychosis in the research field.

This thesis reports the results of a systematic review of the preclinical literature of psychotic disorders wherein studies that test animal models of psychotic disorders against naïve, control animals have been identified to help our understanding of the underlying pathophysiology of these conditions in humans. In addition, I also include studies that investigate the therapeutic potential of certain treatment drugs. I review the literature in the context of four themes. I explore whether limited translation might be explained by a) a disorganized field whereby the identification of patterns in information becomes difficult, b) experiments carried out and published being of limited relevance to the clinical condition, c) studies published and methodological designs of experiments being of limited quality, and d) the data that can be captured from the literature being skewed or misrepresentative of the field.

Chapter 2 describes the methodology I have used for this review, including explanations of calculations for meta-analyses carried out. Results of the search and the characterization of studies identified is presented and discussed in Chapter 3. Here I include a summary of models most commonly used, outcomes most widely measured and a list of treatment compounds that are tested in the literature identified. In Chapter 4 I take a subset of this data forward to explore four forms of validity – external, face, construct and predictive – in order to assess the relevance of animal models and experimental paradigms describing these have for the clinical condition. In Chapter 5 I introduce a novel technique for the categorisation of publications according to the reporting of a list of risk of bias and other methodological quality criteria. This is an automated tool that was developed while spending time at Óbuda University in Budapest, Hungary and uses text mining to speed up the process of reading and extracting data from publications. In Chapter 6 I demonstrate the use for this tool, while reviewing the reporting of risk of bias items and other methodological quality criteria across the entire field. In Chapter 7 I

explore the difficulty in capturing all of the relevant data in relation to my research question and thus review the robustness of my obtained data. Finally, Chapter 8 summarizes my results and thoughts on the field, discussing any trends identified and consequent suggestions I believe might help advance our understanding further of the use of animal models of psychotic disorders.

The thesis will focus mainly on schizophrenia as this is the most commonly described psychotic disorder in the field of preclinical research. Nevertheless, as psychotic symptoms are shared by all of the psychotic disorders, any data reporting outcomes measured in animals thought to be of relevance to human psychosis is likely to improve our understanding of all these disorders and I try to make inferences to all psychotic disorders where possible. As a result this review includes all outcomes described in the context of psychotic disorders including negative symptoms and cognitive deficits.

I believe that the data collected here is beneficial to the field because it is becoming more and more apparent that there is no panacea for all individuals who experience psychotic symptoms. It is imperative that we create a well-structured system of preclinical research that is continuously informed by and built on existing knowledge. For this, research needs to be comprehensive and informative, exploring different methodological designs to define the best dose, route and frequency to take forward to clinical trials; complete, so that all research endeavours are made public; and readily translatable so that animal models more closely model human pathogenesis and thus have maximised value in the drug discovery process for human psychotic disorders. The field of pre-clinical psychotic research has evolved over the years with our ever-increasing knowledge, and this knowledge has thus increased the breadth of research that is currently being carried out. There is a risk, however, that this will also potentially lead to a large number of false positive if research is not sufficiently guided by theories of aetiology, pathogenesis and cognition in the context of human psychosis (Moore, 2010). A more detailed and structured roadmap of the current field, would allow future studies to identify which models work best for modelling different clusters of signs resembling human symptoms of psychosis and therefore identify treatment drugs that might give equally promising results in the clinic for the management of psychotic disorders.

2 Methods

In this chapter, I discuss the general methodology that I used to conduct my systematic search of the research field and how I extracted the data I used for meta-analysis at later stages of the project. Any other methods used throughout my project, more specific, are described within the corresponding chapter. I have also developed a new methodological technique to trial the use of text mining in order to speed up the process of reviewing the studies I included in my review, and this is discussed in Chapter 5. As a recognition of the importance of transparency and in light of the importance of specifying your methodology at the start of a project to avoid the introduction of bias, a detailed outline of the methods used for the main systematic review and meta-analysis part of the project has been published in a protocol format (Bahor et al., 2016). The contribution of other individuals to my work is acknowledged in more detail at the beginning of each specific chapter. Within this chapter I refer to Cristina Nunes-Fonseca (CF) and Hannah Vesterinen (HV) for their work.

2.1 Systematic Review

Studies that characterise animal models of psychotic disorders and/or test interventions in these animal models were identified from the literature using the electronic database PubMed. I chose to search a single database to perform a shallow, but broad review of the field, and thus get an exploratory snapshot of the research that was out there. This of course could have introduced a certain extent of selection bias into my overall findings and affected the robustness of my overall results (Higgins and Green, 2008), which is a limitation of the current work. While much of the literature on clinical systematic reviews recommends performing systematic review searches using multiple electronic databases (Higgins and Green, 2008; Stevinson and Lawlor, 2004), there is also some evidence that has found performing these additional searches only has a modest impact on overall results (Halladay et al., 2015), with little change to overall statistical significance of results (Hartling et al., 2016). Guidance for systematic reviews of animal models has so far been less clear-cut and certainly the effect of how many and which databases are searched seems highly dependent on the topic being explored even for human studies (Hartling et al., 2016).

I defined psychotic disorders as a group of psychiatric disorders including non-affective psychotic disorders, affective psychotic disorders, substance-induced psychotic disorders and psychotic disorders due to a general medical condition. Studies were identified using search terms synonymous to the word “psychosis” and a list of psychotic mental disorders. These disorders were primarily decided on using the classification list of mental disorders outlined in Chapter V of the International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) (World Health Organization, 2016). Information from the DSM-5 Schizophrenia Spectrum and Other Psychotic Disorders chapter (American Psychiatric Association, 2013) was also used, to include only those that I considered to be primarily psychotic disorders (Figure 2.1).

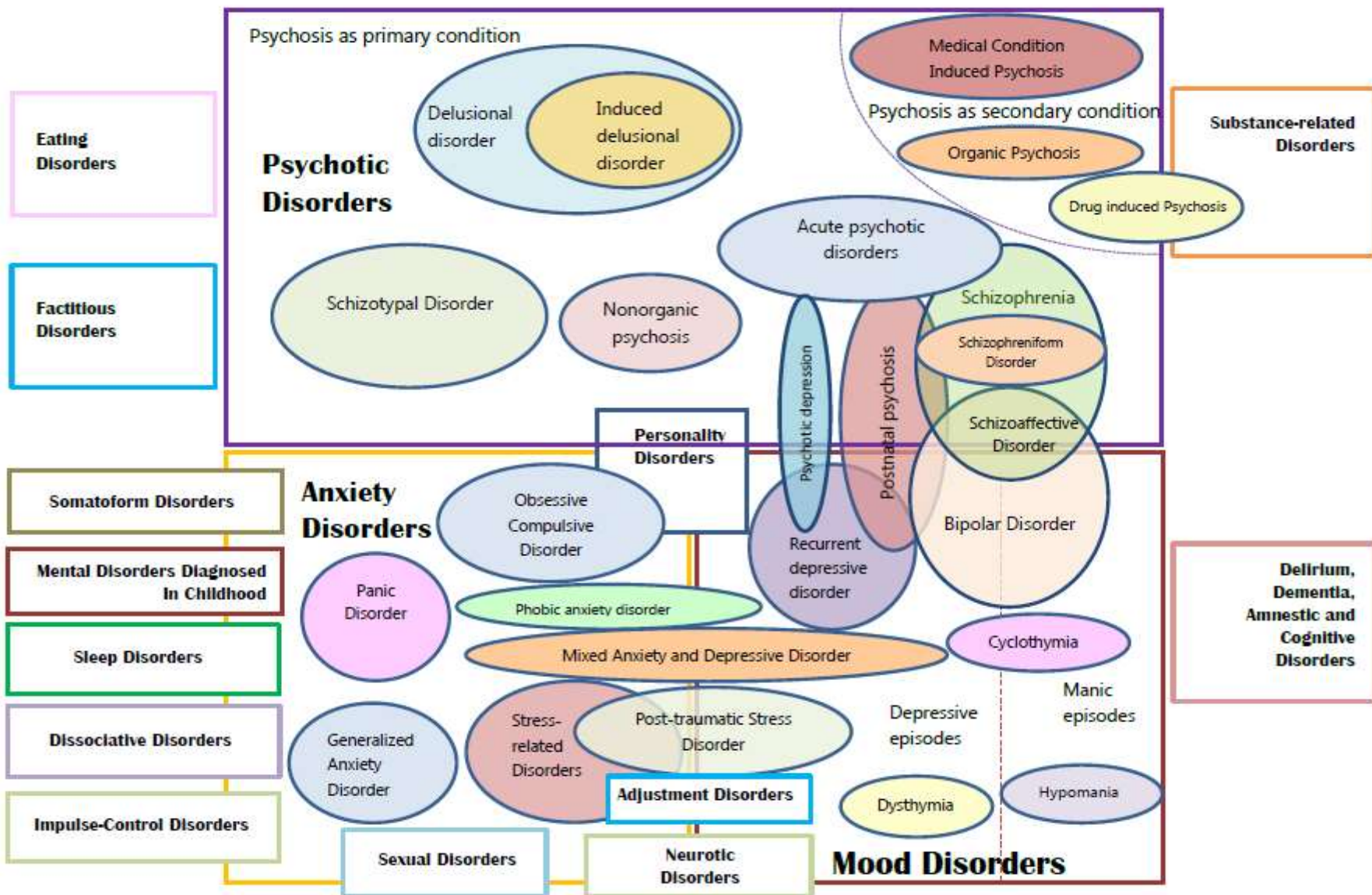


Figure 2.1. Venn Diagram of Disorder Classifications based on ICD-10 and DSM-5 for three major groups of disorders –psychotic, mood and anxiety disorders.

Using ICD-10, psychotic disorders were identified under Chapter 5 of Mental and Behavioural Disorders, which included disorders of psychological development. More specifically, I chose to include the following items in the list of classifications: Schizophrenia, schizotypal and delusional disorders (F20-29), which covered schizophrenia, schizotypal disorder, persistent delusional disorders, induced delusional disorders, schizoaffective disorders, acute and transient psychotic disorders, other nonorganic psychotic disorders and unspecified nonorganic psychosis. This list was supplemented using the chapter Schizophrenia Spectrum and Other Psychotic Disorders from the DSM-5 to also include catatonia, substance- or medication-induced psychotic disorder and psychotic disorders due to another medical condition. I also considered affective disorders with psychotic features to be relevant in this review. Based on my list of disorders identified through the two classification systems used, and associated synonyms of these disorders I performed the search using the terms:

((((((((((((((((((((((((((psychot*) OR psychosis) OR psychoses) OR paranoia) OR paraphrenia) OR sensitive beziehungswahn) OR involuntional paranoid state) OR folie deux) OR cataton*) OR delusion*) OR hallucinat*) OR schizotyp*) OR psychoactive) OR oneirophrenia) OR psychogen*) OR bouffee delirante) OR hebephrenia) OR schizophren*) OR schizoaffect*) OR manic stupor)) NOT comment*[Publication Type]) NOT case report*[Publication Type]) NOT letter*[Publication Type]) NOT review*[Publication Type];

In light of the project's exploratory nature, I tried to keep search terms as broad as possible to maximise the number of relevant publications captured. For this same reason, I did not include any outcome measure terms in the search criteria. Other than including exclusions for certain publication types within the search terms, the search was also limited by an animal filter published previously in the literature by Hooijmans *et al.* (Hooijmans *et al.*, 2010b), claiming to be an improved alternative to the default animal filter available on PubMed already. The search was completed on January 2013. No limits were imposed on this identified set of studies in terms of date published or language published in. A more detailed and expanded list of search terms can be found in the Appendix.

The results of the search were exported into EndNote 6.0, which was used to also import any available PDFs through the University's subscription system. Then using Reference Manager 12, studies were imported into an MS Access database I built to

be able to review studies in more detail during the study selection stages. The architecture of this database was simple, in that all the data were held within one table where each record corresponded to a single study identified within my search and held bibliographic information about that study. I then designed forms for each of the selection phases of the review process, which was essentially a user-friendly interface that filtered the relevant fields of each record to display only those that were required for that phase of the project.

2.1.1 Stages of study selection and data extraction

Study selection of relevant studies for this review was divided into three main phases that coincided with two levels of data extraction:

- (1) Pre-screening of studies based on title and abstract to identify relevant studies and exclude non-primary animal studies;
- (2) Full text screening: involving categorizing publications according to model used, treatment tested and outcome measured; and the recording of reporting of experimental risk of bias; and
- (3) Screening at the level of data extraction for papers looking at behavioural outcome measures.

The extraction of information regarding the reporting of experimental risk of bias items was simply to obtain an overall picture of the extent that these measures are reported in the literature and I did not use this as a variable to exclude papers from further analysis.

2.1.2 Phase I: Pre-screening based on title and abstract

After identification of studies, the first shallow level of inclusion and exclusion of studies was performed, based on the title and abstracts of all identified papers (Figure 2.2). Studies were screened by two independent reviewers (myself and CF) against a pre-defined inclusion and exclusion criteria, and a third independent investigator (HV) resolved any disagreements. During this first phase of the screening process we categorized publications by the human condition they claimed to model. Categories were schizophrenia, substance-induced psychosis, medical condition-induced psychosis, postpartum psychosis and unspecified psychosis.

2.1.2.1 Inclusion/Exclusion criteria

This criteria demanded the inclusion of studies that reported experiments describing an *in vivo* model of a psychotic disorder and measuring an anatomical, electrophysiological, neurochemical, or psychosis-related behavioural outcome, and/or tested the effect of therapeutic interventions on the same outcomes in these models. This meant that studies characterising an animal model where these animals were compared to healthy animals; and those testing the effect of a drug in some of these models and comparing them to untreated animals, were included. Overall, studies were retained if they were primary research articles involving whole, non-human animal models of psychotic disorders that intended to model any human symptom related to psychotic disorders specified above, including transgenic animal models. This meant that any studies that were case reports, human studies, letters or comments, reviews, conference and seminar abstracts without data were excluded. Only studies looking at non-human animals were retained and all other experiments including those performed on *in vitro* samples were excluded. Abstracts were assessed for the reporting of an induction of any psychosis-related or other schizophrenia-associated behaviours or structural changes in the animal described. At this stage studies that did not report any of the outcome measures deemed appropriate for this review (behavioural, anatomical, electrophysiological and neurochemical) were also excluded from further analysis. Experiments considered to be of relevance were not limited in any way to animals of a specific age, sex or species; or interventions of a specific dosage, duration or frequency of administration. Studies of *in vivo* experiments where no disease model had been induced before treatment administration (i.e. drugs were not given to alleviate phenotype of model, such as toxicity studies examining side-effects of anti-psychotic drugs) were excluded as these experiments were not recognised to represent animal models of psychotic disorders. Moreover experiments investigating drug withdrawal, drug discrimination and any other drug addiction investigation models in animals, were excluded. All publications that had not been excluded at this stage by this pre-defined inclusion and exclusion criteria went on to the next phase of the project (Figure 2.2).

S2 Psychosis Screener Form

Psychosis 2

ID 49C RefMan ID 488 Year 2004 281 - 299

Title Co-expression and in vivo interaction of serotonin1A and serotonin2A receptors in pyramidal neurons of prefrontal cortex

Abstract: The prefrontal cortex plays a key role in the control of higher brain functions and is involved in the pathophysiology and treatment of schizophrenia. Here we report that approximately 60% of the neurons in rat and mouse prefrontal cortex express 5-HT(1A) and/or 5-HT2A receptor mRNAs, which are highly co-localized (approximately 80%). The electrical stimulation of the dorsal and median raphe nuclei elicited 5-HT1A-mediated inhibitions and 5-HT2A-mediated excitations in identified pyramidal neurons recorded extracellularly in rat medial prefrontal cortex (mPFC). Opposite responses in the same pyramidal neuron could be evoked by stimulating the raphe nuclei at different coordinates, suggesting a precise connectivity between 5-HT neuronal subgroups and 5-HT1A and 5-HT2A receptors in pyramidal neurons. Microdialysis experiments showed that the increase in local 5-HT release evoked by the activation of 5-HT2A receptors in mPFC by DOI (5-HT2A/2C receptor agonist) was reversed by co-perfusion of 5-HT1A agonists. This inhibitory effect was antagonized by WAY-100635 and the prior inactivation of 5-HT1A receptors in rats and was absent in mice lacking 5-HT1A receptors. These observations help to clarify the interactions between the mPFC and the raphe nuclei, two key areas in psychiatric illnesses and improve our understanding of the action of atypical antipsychotics, acting through these 5-HT receptors

S2 Included
S2 Excluded
Anti-psychotic testing

Journal Cerebral cortex (New York, NY: 1991) Issue 3 ISSN/ISBN 1047-3211 (Print) 10

PDF Mental Illness\Psychosis publications\488_2004.pdf

Model 1 Induction Treatment given
Model 2 Induction

Author Amargos-Bosch M;Bortolozzi A;Puig MV;Serrats J;Adell A;Celada P;Toth M;Mengod G;Artigas F; Conference title Cereb Cortex

Keywords Amphetamines/pharmacology Animals Autoradiography Dopamine/physiology Electric Stimulation In Situ Hybridization Male Mice Mice, Inbred C57BL Mice, Knockout Field36 <http://cercor.oxfordjournals.org/content/14/3/281.full.pdf>

Details engAmargos-Bosch, Merce Bortolozzi, Analia Puig, M Victoria Serrats, Jordi Adell, Albert Celada, Pau Toth, Miklos Mengod, Guadalupe Artigas, Francesc Research Support, Non- Field37 <http://cercor.oxfordjournals.org/content/14/3/281.full.pdf>

Field34 Field19

Record: 490 of 14721 No Filter Search

Figure 2.2. Screenshot of the form used in MS Access in order to screen studies identified in search against inclusion/exclusion criteria. There was a separate form for screeners 1 and 2 that looked identical.

2.1.2.2 Disagreements between reviewers

After phase I screening of abstracts by two screeners, a query was created to filter out and identify studies where there were differences in the screening classifications made by these screeners. These were reconciled by a third investigator, using separate forms created especially to display these results where studies could also be categorized based on full-text screening (Figure 2.3). As this was going on at the same time as Phase II of the project, forms for these stages were very similar.

Figure 2.3. Form used for reconciliation of studies not agreed upon by screeners 1 and 2 regarding inclusion/exclusion as well as simultaneous categorization.

2.1.3 Phase II. Full text screening, categorization and assessment of reporting of experimental risk of bias

The second phase of study selection was the first stage in the review process where studies were screened on a full-text basis (Figure 4). This phase of the screening process involved a detailed categorization process and was also the stage at which the reporting of measures taken within these experiments to reduce the risk of bias was assessed for. Unfortunately, where I was unable to obtain the PDF for a full-text article, the publication had to be excluded from further analysis as an abstract was simply not considered reliable enough to be able to categorise and ascertain reported study quality from.

2.1.3.1 Inclusion/Exclusion criteria

Here those studies that did not report the induction of a whole animal model of a psychotic disorder, including studies where a drug considered to be therapeutic for

psychotic disorders was administered without model induction, were excluded. Reviews or otherwise studies with no quantitative data were also excluded from further analysis. Where an experiment measured outcomes in animal models not thought to be of relevance to psychotic disorders (i.e. positive symptoms, negative symptoms or cognitive deficits), it was not considered to be relevant for the current review and was not included in any further stages of the review. Studies only reporting metabolic activity or modelling drug addiction were also excluded at this stage of the review.

Publications that did not report behavioural measures were included still at this phase of the review, but were excluded from any further stages of the review due to time constraints. No studies were excluded based on their level of reporting of measures taken to reduce the risk of bias. Any studies that had to be excluded at this point due to inclusion/exclusion criteria were labelled with a reason for exclusion.

Figure 2.4. Phase II categorization form for studies that had been included by both screeners allowing the input of fields concerning reported study quality, disease model induction details, details of any treatments tested in these models and outcomes measured within the study for these models.

2.1.3.2 Categorization

Studies were categorized according to certain aspects of experimental design reported. Here details about the exact method used to create the animal model, the names of any treatments tested in these models and finally the type of outcome measure recorded during experimentation were logged. More specifically, I recorded information about the method used within a study to induce the experimental condition used to model the disorder of interest in a model animal. Where more than one model was used, I recorded each relevant model mentioned. Where an animal was given two different hits to model a condition, this was recorded as an example of a combination model. Where comparisons for both single and combination models could be identified, both were recorded (Figure 2.5).

The screenshot shows a web application interface for study quality assessment. The main title is "Both Included Study Quality". The interface is divided into several sections:

- Form Fields:** ID (1324), RefMan ID (1323), Title (Action of amphetamine and some psychotropic agents on the acquisition of an escape and avoidance conditioning in rats selected as a function of the especially low level of their...), PDF (Mental Illness\8_publications\1323_1965.pdf).
- Buttons:** "NEXT", "Add New Reason", "Add New Outcome Measure", "OK", "Cancel".
- Study Quality Section:** Includes checkboxes for Randomisation, Blinded conduct of Experiment, Blinded assessment of Outcome, Inclusion/Exclusion criteria, Sample Size Calculation, Conflict of Interest Statement, Compliance with Animal Welfare Reg, and Study Protocol Availability.
- Model Specifics Section:** Includes a dropdown menu for "SPECIFIC MODEL" with a list of models, and checkboxes for "Combination Model" and "Treatment Tested".
- Outcome Specifics Section:** Includes checkboxes for "Behavioural outcome", "Locomotor activity measured using computer?", "Anatomical outcome", "Electrophysiological outcome", "Neurochemical outcome", and "Outcomes in Same Cohort".
- Other Fields:** "Categorised/No RoB" checkbox, "Reason for exclusion" dropdown (Need to translate (french)), "Complete (outcomes)" checkbox, and a date field (01/05/2017).

Figure 2.5 Screenshot of drop-down options for studies where more than one model was reported within the same study

Experimental designs where treatments were administered to animal models, were also recorded in a separate field. Where studies were only looking at characterising the model and were given no therapeutic interventions this field was left blank. I recorded all treatments tested in relevant models even if they were reported to not have any significant effect on the model.

In a similar way, I recorded the outcome measures that were reported within each study. Specific detail such as the exact name of outcome measure and what it measured was only recorded for behavioural outcome measures as these were the ones data was extracted for in the next phase of the project.

Categorization of studies was not only done to get a better overview of the field, but also in order to allow for the filtering of studies at a later time point for specific models. For studies that needed to be reconciled from disagreements at phase I of the review, studies were categorised in a similar way. This was carried out jointly by myself and HV.

2.1.3.3 Extraction of details about reported methodological quality and reported risk of bias

During this phase of categorization, studies were also scored against an 8-item study design checklist. This was to assess for both the extent to which bias might have been introduced during experimentation as reported within a study, as well as a possible reduction in imprecision and quality due to the lack of reporting of certain methodological aspects.

This built on a checklist used by our group in previous projects (Macleod *et al.*, 2004), and was adapted to include the following items: (1) Randomisation; (2) Blinded conduction of experiment; (3) Blinded assessment of outcome; (4) Statement of Inclusion and Exclusion Criteria; (5) Sample Size Calculation; (6) Statement of possible Conflict of Interest; (7) Statement of compliance with Animal Welfare Regulations; and (8) Availability of a study protocol.

I was interested in the prevalence of reporting of each item and study characteristics that might be related to these results (Figure 2.6). Two independent investigators normally carry out this phase of a systematic review project, however, resources only allowed for screening by a single investigator for each study here. An automated technique was developed (see Chapter 5) to address difficulties with the screening of so many studies. I did not exclude any studies from my final analysis, based on their overall study quality score as I wanted to assess the impact of these factors on the outcomes that were reported in the studies analysed.

Risk of Bias		Method
Study Protocol Availability	<input checked="" type="checkbox"/>	<input type="text"/>
Random Allocation to Group	<input checked="" type="checkbox"/>	<input type="text"/>
Allocation Concealment	<input type="checkbox"/>	<input type="text"/>
Blinded Assessment of Outcome	<input type="checkbox"/>	<input type="text"/>
Sample Size Calculation Performed	<input type="checkbox"/>	<input type="text"/>
Compliance with Animal Welfare Regulations	<input checked="" type="checkbox"/>	<input type="text"/>
Statement of Potential Conflicts of Interest	<input checked="" type="checkbox"/>	<input type="text"/>
Reporting of animal exclusions	<input type="checkbox"/>	Numb Reporte Pre-defi

Figure 2.6. Screenshot from the database used to record details about the reporting of items from our list of study quality items.

2.1.4 Phase III: Collection and Extraction of Data from Publications

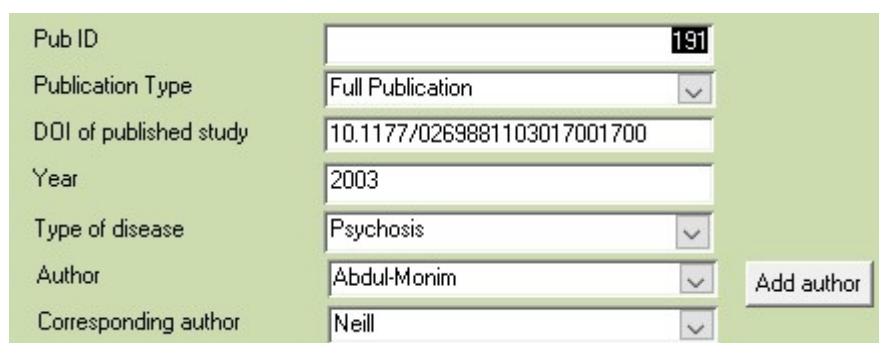
The final phase of screening was the final inclusion for data extraction and meta-analysis and involved the screening of studies on a full-article level. At this phase of screening an in-depth level of data extraction was performed, where further details about the experiments performed and outcome data reported within each paper were recorded and extracted.

2.1.4.1 Inclusion/Exclusion criteria

As mentioned before, this was the stage I excluded studies not reporting a behavioural outcome from further analysis. I also excluded studies where data being referred to was not clear from a publication or were missing appropriate control groups. For meta-analysis three important pieces of data (n numbers, mean, SD/SEM) are required, therefore I could not include studies where this was not clearly reported in any further stages of this review. Any studies not meeting these criteria were excluded from the meta-analysis, but were retained for an overall review of the field (i.e. results from Phase II of the project) that describes this particular research field of interest.

2.1.4.2 Extraction of details about publication and model used

Study characteristics that were extracted included further details specific to each study: name of first and corresponding authors, year of publication, title of article and name of journal. I was also interested in recording the type of journal article data were extracted from, to see whether this could have an effect on the reporting of certain kinds of data (Figure 2.7).



The screenshot shows a form with the following fields and values:

Pub ID	191
Publication Type	Full Publication
DOI of published study	10.1177/0269881103017001700
Year	2003
Type of disease	Psychosis
Author	Abdul-Monim
Corresponding author	Neill

An "Add author" button is located to the right of the Author field.

Figure 2.7. Screenshot from the database used to record study identification details.

Details about the animal model used within an experiment were also collected and recorded to include species, strain and sex, labelled as „unknown” if otherwise not specified. Age at the time of testing and weight of animals used was also recorded if it was reported, if not these fields were left blank. I also chose to record specific details about the method of model induction including type of model the model could be categorized as – this would be one of the following: pharmacological, developmental, lesion, genetic or a combination model. Other data that were extracted in this section were: details about the specific method used to induce the model; the site of this induction or injury; the dose or severity of the damage, where appropriate; details of the time and duration of administration, damage or exposure to condition; and control animals used to compare experimental groups to (Figure 2.8).

Treatment Drug, if given (Otherwise *)	Haloperidol	<input type="button" value="Enter New Drug"/>
Species	Rat	
	Hooded Lister	
Animal Supplier and Country	Harlan	UK
Sex	Female	
Neuropathic Pain Model Details		
Anaesthetic used for surgery		
Type of Psychosis Model	Psychosis Developmental Model	
Specific Method of Model Induction	Isolation Rearing	
	<input type="button" value="Add method of induction of injury"/>	
Site of Induction	Intraperitoneal	
Dose/severity to induce model	2 mg/kg	
Total number of events/administrations	0	
Number of events/administrations (per day)		
Day of last event/administration		
Control used	Other	
Duration of Isolation (days)		
Weight range	250	300
Age range (weeks) or categorical age		Adult

Figure 2.8. Screenshot from database detailing fields where data about animal model characteristics were entered.

For developmental studies only (i.e. such as those inducing the animal model using isolation rearing), where other details such as husbandry and housing were important, additional fields were used to extract these data. Data were taken, where reported, detailing the conditions that experimental animals were placed in, such as diet, food availability, details about housing and conditions of laboratory used to keep animals as well as handling of animals (Figure 2.9). These were items that were based on a previous systematic review of animal models of neuropathic pain, where data about husbandry of animals were taken in greater detail (Seretny et al., 2014).

Isolation Details

Diet	Standard	▼
Food availability	Restricted availabil	▼
Bedding	Unknown	▼
Cage rack ventilation	<input checked="" type="checkbox"/>	
<u>Home cage enrichment:</u>	-Shelter	<input checked="" type="checkbox"/>
	-Nesting material	<input checked="" type="checkbox"/>
	-Chew block	<input checked="" type="checkbox"/>
	-Other	<input type="checkbox"/>
Number housed per cage range in control group	5	
Control and Model animals in same room	Not Mixed	▼
Room temperature (°C)		21
Humidity (%)		40
Frequency of handling by experimenters		0
Cage cleaning frequency per week:		0
Different species housed in the same room	<input checked="" type="checkbox"/>	
Habituation time pre-study (days)		0

Figure 2.9. Screenshot of database showing fields used to extract data about animal husbandry and housing for specific studies

Multiple entries were created for studies where more than one model or more than one treatment drug had been tested. Data for each of these experiments were entered on different pages, but linked by the publication's unique identifier.

2.1.4.3 Extraction of outcome data

All relevant comparisons of behavioural outcomes meeting the pre-specified inclusion and exclusion criteria were identified within each study. This meant that data were extracted separately for comparisons where variables such as age, sex and species of animals differed in the groups of animals being tested in an experiment. Data were also extracted separately for outcomes measured as a result of different treatment regimes. As many of the outcomes reported were measured over a period of time, where data were reported at multiple time points within the same group of animals over time, every single time point was taken, which would later be used to combine and calculate overall performance.

Characteristics of therapeutic interventions tested were extracted for studies that had looked at the effect of a treatment drug. These included the dose of the drug given, mode and time of delivery, frequency of administrations and the length of the treatment regimen. Details about the time taken between the recording of outcome measurements and treatment administration as well as time of outcome assessment relative to time of model induction, were also recorded (Figure 2.10). For studies characterising the model and not testing the effect of a treatment drug these fields were left blank.

Treatment Drug Dose and Units	Route of Delivery	Time of first admin relative to model induction (hours)	Time of last admin relative to model induction (hours)	Total number of admins	Time of outcome assessment (hours)	Age of animals at separation (d) Lower Upper	Time of outcome assessment relative to model (hours)	User Defined 2 User Defined 2a	Time of outcome assessment relative to last drug admin (hours)
0.5 mg/kg	IPeritoneal	24	24	1	1	21	48		0.5
1 mg/kg	IPeritoneal	96	96	1	1	21	120		0.5

Figure 2.10. Screenshot of database where details about treatment tested and time of outcome assessment were recorded.

In terms of outcome measurement details, the exact name of the outcome being measured, the method used to do this, or in other words the tools and way of measurement (e.g. count of beam breaks for locomotor activity), the units of measurement and whether the larger the numerical data recorded would indicate a better or worse outcome, were recorded. For later analysis, a note was also made of the number of experimental groups that were compared to the same cohort of control animals within an experiment, and for multiple measurements taken within the same experiment from the same cohort of animals, letters of the alphabet were used to identify different cohorts of animals (Figure 2.11). These data were important in order to avoid counting the same group of animals more than once at later stages of analysis when it came to pooling the data using meta-analysis.

Outcome Measure	Units of Outcome Measure	Specific Method of Outcome Measurement	Larger value indicates improvement or worsening in outcome?	Number of groups served by this control group
Locomotor Activity	number	LMA count	worse	6
Please tick when entry is complete		<input checked="" type="checkbox"/>	Assigned Cohort Letter	a
Reversal Learning Task	%	Correct responses	better	6
Please tick when entry is complete		<input type="checkbox"/>	Assigned Cohort Letter	b

Figure 2.11. Screenshot of database showing fields where details of outcome measures would be recorded.

For outcome measures of behaviour that required further information, this was extracted as additional information into a 'user defined field' (Figure 2.10). This was utilised mainly for experiments where pre-pulse inhibition (PPI) was measured, and we recorded details such as the strength of pre-pulse stimulus used, that were not otherwise a variable for other behavioural outcome measures.

Within each of these comparisons for each outcome measure a mean, standard deviation or standard error of the mean and number of animals contributing to that mean (n) were extracted for both experimental and control groups. Where n numbers were given as a range, the most conservative estimate was taken (Figure 2.12).

Animal Groups	n	Mean	SEM	SD
				Please enter SEM OR SD
Model+Vehicle	12	807	47	
Model+Treatment Drug				
Control+Vehicle	12	487	31	
Control+Treatment Drug				
Naive Animal				
Model+Vehicle	6	78	4	
Model+Treatment Drug	6	81	10	
Control+Vehicle	6	84	4	
Control+Treatment Drug	6	77	4	
Naive Animal				

Figure 2.12. Screenshot of database showing fields where specific outcome data measures from experiments were recorded.

Preferably, these numerical data were extracted from the text of each publication, including cases where this was presented in a tabular format. In most cases, however, as results were presented graphically, Adobe Reader Measuring Tool in Adobe Acrobat XI was used to obtain numerical data needed for each comparison. It was decided early on that comparisons where data were unclear from graphs, authors would be contacted to clarify values and obtain correct data. In the absence of a response from these authors, data were excluded from further analysis.

Unfortunately, where any of these data (mean, SD or SEM, *n*) were absent in a paper, that specific comparison had to be excluded from further analysis as it could not be included in the meta-analysis, however, these papers were retained in the overall descriptive summary of the systematic review.

2.2 Meta-analysis

I describe my methodological approach to the meta-analysis part of the project, based on the stages that normally follow the process of a meta-analysis. First, for each comparison, I calculated an effect size, which were then weighted and used to calculate a summary effect size. Finally, heterogeneity present in the data were calculated, to explore the impact of pre-defined study characteristics and how much of the heterogeneity can be attributed to any of these variables (Vesterinen et al., 2014).

2.2.1 Identification and definition of comparisons

Studies characterising the model and those testing the effect of a drug in these models, were chosen to be analysed separately. Based on this a comparison was defined as a measure of neurobehavioural outcome in a group of animals that had been exposed to an intervention believed to induce an animal model of a human psychotic disorder, compared with a group of animals that had not been exposed to this intervention. Suitable controls in these comparisons were considered to be naïve animals for developmental and genetic studies, those that had been given sham surgery in the place of the lesion for lesion studies, and the administration of saline or another vehicle for pharmacological studies. Comparisons for studies testing the efficacy of a therapeutic intervention, were defined as a comparison of neurobehavioural outcomes measured in two different groups of animals that had been exposed to the same intervention to induce the model, but one had also been given a treatment drug to ameliorate behaviours, while the control group had not. Suitable control groups in this case for pharmacological studies included both those animals that had been given a vehicle as a treatment and non-treated animals.

For analysis where data were reported from separate groups of animals, data collected from these groups were treated as independent comparisons and were included separately within the meta-analysis. Where more than one outcome measure was measured and reported for the same group of animals within a study, data were nested using the fixed effects model. This was considered separately for model-characterising comparisons and therapeutic drug comparisons. Where a

control group served multiple experimental groups within a study, the number of animals contributing to the meta-analysis was corrected for by dividing the number of animals that was reported by the number of experimental groups it had served.

2.2.2 Individual comparison effect size estimate

The data being extracted from publications included in the review were continuous data. These data required the extraction of a mean outcome of all animals in each group and its variance. Using these data it would have been possible to calculate an absolute difference in means, a normalised difference in means, or a standardised difference in means.

An absolute difference in means (MD) allows for the simple calculation of the difference between the reported means of the two groups (control and treatment) within a comparison and its corresponding variance (equations 2.1, 2.2, and 2.3)

$$MD_i = \bar{x}_c - \bar{x}_{Rx} \quad 2.1$$

Where

$$\bar{x}_c = \text{sample mean of control group animals}$$

$$\bar{x}_{Rx} = \text{sample mean of treatment group animals}$$

With Standard error calculated as:

$$SE_i = \sqrt{\frac{N}{n_{Rx} \times n'_c} S_{pooled}^2} \quad 2.2$$

Where

$$n_{Rx} = \text{number of animals in treatment group}$$

$$n'_c = \text{true number of control animals (corrected for number of treatment groups served by same control)}$$

$$S_{pooled} = \text{the pooled standard deviation calculated}$$

$$N = \text{total number of animals in treatment group and in adjusted control group}$$

With S_{pooled} calculated as:

$$S_{pooled} = \sqrt{\frac{(n'_c-1)SD_c^2 + (n_{Rx}-1)SD_{Rx}^2}{N-2}} \quad 2.3$$

Where

SD_c^2 = reported standard deviations for control group (convert from standard error if necessary)

SD_{Rx}^2 = reported standard deviations for treatment group (convert from standard error if necessary)

I decided against using this method to estimate effect size for each comparison, because this method would have required my data being analysed to have come from the same outcome measure and be reported on the same scale across all my included studies. Considering the large differences in outcomes measured and the method used to do this across different studies, this would not have given a reliable effect size estimate.

Normalised mean difference (NMD), an alternative method, is appropriate for data where the score of a normal, untreated or “sham” animal is already known or can be derived and therefore we can use this to express an absolute difference in means as a proportion of the mean in the control group (equations 2.4, 2.5 and 2.6)

$$NMD_i = \frac{(\bar{x}_c - \bar{x}_{sham}) - (\bar{x}_{Rx} - \bar{x}_{sham})}{(\bar{x}_c - \bar{x}_{sham})} \times 100\% \quad 2.4$$

Where

\bar{x}_c = sample mean of control group animals

\bar{x}_{Rx} = sample mean of treatment group animals

\bar{x}_{sham} = sample mean of sham group

Standard error calculated as:

$$SE_i = \sqrt{\frac{SD_{C*}^2}{n_C} + \frac{SD_{Rx*}^2}{n_{Rx}}} \quad 2.5$$

Where

SD_{C*}^2 = standard deviation of control group expressed as a percentage of the control group and normalised to the sham group value

SD_{Rx*}^2 = standard deviation of treatment group expressed as a percentage of the control group and normalised to the sham group value (see below)

Normalised standard deviations calculated as:

$$SD_{C*} = 100\% \times \frac{SD_C}{\bar{x}_c - \bar{x}_{sham}} \quad \text{and} \quad SD_{Rx*} = 100\% \times \frac{SD_{Rx}}{\bar{x}_c - \bar{x}_{sham}} \quad (2.6)$$

Where

SD_C = reported standard deviation for control group

SD_{Rx} = reported standard deviation for treatment group

It was decided right at the start of the project that where the performance of a naïve, unlesioned or wild-type animal could be inferred or was already known in at least 80% of the experiments, NMD would be used to calculate individual effect size estimates as the primary outcome and standardised mean difference as a sensitivity analysis. If the performance of these animals was not known for more than 80% of the experiments then it was decided I would use NMD only as a sensitivity analysis and instead would use standardised mean difference meta-analysis as the primary outcome.

Standardised mean difference (SMD) is mostly used where we do not know or cannot infer how a naïve animal would perform. This is especially relevant for some of our data where spontaneous locomotor activity would be measured in a group of animals as a neurobehavioural outcome. This method calculates a difference between the means of each group and divides this by a pooled variance. This allows

for all outcome measures to be changed so they now are on a standardised scale with units given in standard deviations (SDs)(equations 2.7 and 2.8).

$$SMD_i = \frac{(\bar{x}_c - \bar{x}_{Rx})}{S_{pooled}} \times \left(1 - \frac{3}{4N-9}\right) \times direction \quad 2.7$$

Where

S_{pooled} = the pooled standard deviation calculated as before

Direction = correction factor used to define the direction of the effect size

Standard error calculated as:

$$SE_i = \sqrt{\frac{N}{n_{Rx} \times n'_C} + \frac{ES_i^2}{2(N-3.94)}} \quad 2.8$$

This approach is helpful for analysis of data where performance for the same outcome reported are on different measurement scales. This is especially important for experiments where locomotor activity is being assessed as this can be recorded by counting number of beam breaks, analysing distance travelled over a period of time or number of runs, *etc.*

Hedge's G was used to calculate standardised mean difference, where appropriate. The reason for using this method was because this equation corrects for bias introduced by small sample sizes. This is especially relevant as animal studies in general and thus many experiments included in this review include small samples of animals in each of their groups ("small sample" defined here as less than 10 animals per group). Importantly though, here we had to take into account the direction of the effect (whether a higher score represents a better or worse outcome) and so this direction of effect was recorded at time of data extraction and later used to adjust overall effect sizes accordingly (i.e. multiply by 1 or -1).

2.2.3 Weighted and pooling of effect sizes to give summary effect size

Since this review combined multiple studies of various precision, a weighted mean was calculated for each comparison, before all the data could be pooled. This was essential for the estimation of the overall effect size as this would allow for the studies that were more precise to carry more weight in their contribution to the final estimated combined effect, than those that were less precise and thus carried less information. The calculation of the distribution of effect sizes from individual comparisons had an estimated summary estimate (the weighted mean), an estimate of heterogeneity (the weighted sum of the deviations from this mean squared) and finally a measurement called tau-squared to represent the estimate of the variation between observed effects across different studies beyond that which we would expect to be explainable by random sampling error.

For computing a pooled effect size there are two models available in meta-analysis – the fixed effects model and the random effects model.

Using the fixed effect model we assume that there is a single true effect size where different studies give an estimate of the same effect. This makes the combined effect an estimate of the common effect size, where individual observed effect sizes deviate from this true effect simply as a result of random sampling error. Here I would have calculated the weight of each comparison using the inverse variance method, where each individual effect size would be multiplied by the inverse of their standard error (SE) squared (equation 2.9).

Weight of each study under the fixed effects model:

$$W_i = \frac{1}{SE_i^2} \quad 2.9$$

Therefore weighted effect size under fixed effects model is:

$$W_i ES_i = ES_i \times \frac{1}{SE_i^2} \quad 2.10$$

Then individual effect sizes are multiplied by their attributed weight (equation 2.10, above), and the sum of these calculations is divided by the sum of the weights alone, to give us a summary effect estimate (equations 2.11, 2.12, and 2.13).

$$ES_{fixed} = \frac{\sum_{i=1}^k W^* ES_{\theta i}}{\sum_{i=1}^k W^*} \quad 2.11$$

Where ES_{θ_i} = effect size estimate for each individual study

W^* = calculated weight for each study, so that

$W^*ES_{\theta_i}$ = weighted effect size estimate for each individual study

With standard error calculated as:

$$SE_{fixed} = \frac{1}{\sqrt{\sum_{i=1}^k W^*}} \quad 2.12$$

And 95% Confidence intervals as:

$$95\% CI = ES_{fixed} \pm 1.95996 \times SE_{fixed} \quad 2.13$$

Under the random effects model we expect the true effect to vary between different studies as a consequence of varying study characteristics used in different experiments (i.e. an array of different drug doses, animal species, or methods of model induction used). Therefore, the combined effect will represent the mean of the population of true effects and will take into consideration both the sampling error we consider in the fixed-effects meta-analysis (within study variance), as well as differences in true effect sizes between studies (between-study variance). For this reason when weighing studies under this model, the inverse of the sum of within study variance has to be used (inverse variance calculation as before) and also Tau-squared (τ^2), which is a measure of excess-between study variation (equations 2.14, 2.15, 2.16, 2.17, 2.18 and 2.19). τ^2 is estimated using the method of moments (DerSimonian and Laird, 1986).

Weight of each study under random effects model:

$$W_{+\tau^2}^* = \frac{1}{(SE_{\theta_i}^2 + \tau^2)} \quad 2.14$$

Where tau is estimated using the equation:

$$\tau^2 = \frac{Q-df}{c} \quad 2.15$$

Where τ^2 = estimation of between-study variance

Q = sum of the squared differences in effect sizes between studies and the pooled effect size (see equation 2.20 further below)

df = degrees of freedom calculated by taking number of components in the strata and deducting 1

C = measure used to convert the heterogeneity value into an average and put the value back into original units

Therefore, weighted effect size under the random effects model:

$$ES_{rand}^* = ES_{\theta i} \times W_{+\tau^2}^* \quad 2.16$$

Once studies have been weighted, I could calculate a random effects estimate of the combined effect (equation 2.13, 2.14, 2.15)

$$ES_{random} = \frac{\sum_{i=1}^k ES_{rand}^*}{\sum_{i=1}^k W_{+\tau^2}^*} \quad 2.17$$

With standard error calculated as:

$$SE_{random} = \frac{1}{\sqrt{\sum_{i=1}^k W_{+\tau^2}^*}} \quad 2.18$$

And 95% Confidence intervals same as before:

$$95\% CI = ES_{random} \pm 1.95996 \times SE_{random} \quad 2.19$$

I used the fixed-effects meta-analysis to combine multiple outcomes from the same experimental cohorts as here performance would be obtained from the same population of animals. For the overall statistical model of analysis when it came to estimating the combined effect, I used the random-effects model of analysis as there was considerable amount of diversity in study design characteristics.

2.2.4 Assessing for heterogeneity

Considering the large amount of heterogeneity across studies included in my review with respect to experimental design, I thought it would be interesting to know whether any differences between groups such as certain study characteristics influenced reported outcome. To assess heterogeneity Cochran's Q and I^2 was calculated.

Cochran's Q is the estimate of between study heterogeneity and is calculated as the weighted sum of differences between individual effect sizes and the pooled effect across studies under the fixed effects model squared (equation 2.20).

$$Q = \sum_{i=1}^k W^* \times (ES_{\theta i} - ES_{fixed})^2 \quad 2.20$$

Where W^* = calculated as before in equation 2.9

The values of Q follow a chi-squared distribution under the assumption that if studies come from the same population of studies measuring the same thing, then any heterogeneity we see is a result of sampling error and Q will equal the degrees of freedom. Q is sensitive to the number of comparisons, so that it has a low power of test of heterogeneity when the number of comparisons is small, and too much power when there are a large number of comparisons.

I also calculated I^2 , a statistic that estimates what proportion of the variation across different studies can be attributable to heterogeneity (i.e. true differences in effect sizes) and not chance (equation 2.21). This value is not sensitive to the number of comparisons used within a meta-analysis. The estimate of this value can help to decide whether fixed or random effects model of meta-analysis should be used. While this is subjective, it was decided based on a previous review (Vesterinen *et al.*, 2015) that the random effects model of meta-analysis would be used where I^2 values were above 50% and thus interpreted as moderate or high heterogeneity.

$$I^2 = \frac{Q-df}{Q} \times 100\% \quad 2.21$$

2.2.5 Exploring sources of heterogeneity

Stratified meta-analysis builds on the assumption that studies which share certain characteristics will be similar in terms of outcome than those which do not share these characteristics. So heterogeneity is partitioned into these groups of studies that are similar and between groups of studies. For each strata I would calculate a random effects size and Cochran's Q, which denotes the heterogeneity. These heterogeneity statistics are then summed and deducted from the total heterogeneity to give an estimate of the remaining heterogeneity that we assume comes from the differences between groups. We can test the significance of differences between Q and the variation we would have expected to see if the studies were from the same population of studies using the *chi squared* distribution (equation 2.22).

$$p = CHIDIST(Q_{global} - sum(Q_{strata}), df) \quad 2.22$$

Where Q_{global} = amount of heterogeneity for the global estimate of effect size

Q_{strata} = amount of heterogeneity within components of the strata

df = degrees of freedom calculated by taking number of components in the strata and deducting 1

Meta-regression, takes this approach further and allows the effects of continuous and categorical characteristics to be investigated, where also more than one study characteristic can be taken into account at a time (i.e. multi-variate analysis). This method can tell us more about the data than stratified meta-analysis as it takes into account both within- and between- study variance. The only limitation is that this method requires a large number of studies for the analysis to be meaningful. Meta-regression is a form of weighted linear regression, where we measure the relationship between a dependent outcome variable (our effect estimate) and one or more independent explanatory variable (study characteristics). Therefore, we need to assign a weight to each study and select an appropriate model (fixed or random effects).

If we can presume that the explanatory variable we are interested in is attributable for all the heterogeneity observed between studies, then a fixed-effects meta-regression is appropriate, as this model does not allow for between study variation. Usually, however, it is more suitable to use random effects meta-regression, which allows for extra heterogeneity among the effects that cannot be explained by covariates and therefore both within study and between study variation.

Where appropriate and data were sufficient (i.e. over 25 comparisons included in the meta-analysis), univariate meta-regression was performed to investigate potential sources of heterogeneity. This was decided on in the context that recent research shows that both univariate and multivariate regression are more reliable at detecting an effect of a variable of interest than stratified meta-analysis – which is shown to have substantial false positive rate with NMD estimates and low statistical power with SMD estimates of effect size (Wang et al., 2018). In the current review I describe heterogeneity using τ^2 (estimation of excess between-study variance), residual I^2 (the percentage of residual variation explained by between-study heterogeneity) and adjusted R^2 (adj R^2 ; the proportion of between-study variance explained by the model). Statistical analyses were performed using code written in R for combining data and Shiny Meta-analysis application (<https://qianying.shinyapps.io/Multi/>) for meta-analysis and meta-regression. I use meta-regression to investigate possible sources of heterogeneity including components of study quality and methodological criteria checklist and study design characteristics. I examined different study characteristic subgroups as potential sources of heterogeneity for model-characterising and treatment exploring studies by univariate analysis (Table 2.1).

Table 2.1 Variables of interest in exploring sources of heterogeneity for model-characterising and treatment exploring studies

MODEL-CHARACTERISING STUDIES	TREATMENT EXPLORING STUDIES
Species of animals used	Species of animals used
Gender of animals used	Gender of animals used
Specific intervention used to induce model	Specific intervention used to induce model
For schizophrenia models: Method of model induction (i.e. developmental, genetic, pharmacological, lesion or combination)	For schizophrenia models: Method of model induction (i.e. developmental, genetic, pharmacological, lesion or combination)
Extent of lesion/dose of drug used to induce model	Extent of lesion/dose of drug used to induce model
Outcome being measured	Outcome being measured
Exact methods used to assess outcome measure	Exact methods used to assess outcome measure
Time of outcome measurement (in relation to model induction)	Time of outcome measurement (in relation to model induction and/or treatment administration)
	Treatment given
	Dose of treatment given
	Time of administration

2.2.6 Bonferroni correction

The Holm-Bonferroni method was used to adjust significant values of p . This was necessary, because there were multiple comparisons when assessing differences between subgroups during the partitioning of studies according to certain study characteristics. Study characteristics were grouped separately for those that could be classed as experimental quality items and those that explored study design items. An adjusted p value was calculated independently within these groups.

2.3 Publication Bias

To see whether the potential unavailability of some research results could have skewed our perception of the research field, three different analyses were performed to assess for publication bias. This would appraise the extent to which the published literature used for the basis of my conclusions, was representative of the population of completed studies. This issue of publication bias widely affects all areas of research including pre-clinical studies (ter Riet et al., 2012), where these issues can lead to an overstatement of effects within studies (Sena et al., 2010).

Risk of publication bias was evaluated using funnel plot assessment and Egger's regression. I then used trim-and-fill analysis using STATA to identify possible missing studies in the literature. These evaluations were conducted independently for each outcome measure using non-nested data, to avoid suppressing any studies during analysis that might have contributed comparisons where different outcome measures were reported from the same cohort of animals. I was aware at the time of performance, that these analyses rarely take into account reasons for asymmetry seen in the literature other than publication bias and can be affected heavily by increasing between-study heterogeneity.

2.3.1 Funnel plotting

Funnel plotting uses a simple scatter plot of the relationship between calculated effect sizes plotted on the X-axis against a study's precision on the Y-axis, to allow visual assessment of the potential presence of publication bias. While in systematic reviews it has been common practice to calculate precision as the inverse of the variance for a study (Vesterinen *et al.*, 2014), in light of recent research that has been published I chose to use $1/\sqrt{n}$ to calculate precision. It has been shown that with the use of SMD when an effect is present within a meta-analysis there is a risk for substantial distortion in the estimation of asymmetry when plotting SMD against standard error. It is suggested that for estimation of publication bias for studies of small sizes such as preclinical studies, a sample size-based precision estimate is more suitable (Zwetsloot *et al.*, 2017).

Funnel plots build on the idea that precision of a study increases as the size of the study increases and as a result we see the smaller studies scattered widely on the bottom of the graph, with the larger studies less widely dispersed towards the top of

the graph. If there is no sign of publication bias a symmetrical, inverted funnel shape is expected to be created by the data points centred on the global estimate of overall estimate. If the graph is asymmetrical it can be concluded that publication bias was likely and the more notable this asymmetry, the more likely that I could expect a significant extent of publication bias (Higgins and Green, 2008).

2.3.2 Egger regression

Egger regression builds on the expectation that publication bias is more likely to affect small studies and these are likely to be of lower quality and show larger intervention effects. If this is not true then the effect sizes in a meta-analysis should vary due to random error more for small studies and less so with bigger studies (Egger et al., 1997). For this approach, I plotted a linear regression of normalized effect estimates (effect sizes divided by their standard error) on the Y-axis against precision ($1/\sqrt{n}$) on the X-axis. Whether the regression line crossed the origin was then visually assessed. Presence of publication bias was concluded if it did not. Results were reviewed with the knowledge that this method has a low power for detecting bias in studies of a small sample size, something that is common in pre-clinical studies (Vesterinen et al., 2014).

2.3.3 Trim and fill

The trim and fill method was used to both identify and to correct for potentially missing studies in the literature, using funnel plot asymmetry arising from publication bias. The method first 'trims' small studies from the funnel plot that cause the asymmetry. Secondly, using the funnel plot of remaining studies, it estimates a 'true' centre of the funnel. Finally it replaces the trimmed studies and fills in areas around the centre where potential studies could be missing. This therefore not only gave me an estimate of the number of potentially missing studies, but it also gave me an adjusted effect size estimate, by including filled in studies in a meta-analysis.

3 Systematic Search Results and Characterization of the Literature

3.1 Introduction

Before we can begin to improve something, there needs to be some level of clarity over what that something is and what it is aiming to do. It is very difficult to solve a problem if you do not understand what the problem is in the first place. In order to see whether changes are making a difference we have to quantify the field somehow.

The motivation behind performing a systematic search was the desire to summarise the field of pre-clinical psychosis research in a way that allows for an unbiased, categorized and complete collection of the published literature. The search performed was a shallow, but broad overview aiming to quantify animal studies related to psychosis and psychotic conditions.

First, I wanted to categorise studies according to the experimental design used to investigate these conditions in animals by looking at the way they induced their models, how they quantified these models and if they had tried to ameliorate the effects created in these models at all using an additional intervention. This catalogue of the literature in theory would give us a snapshot of the literature as it stands. I then wanted to explore this catalogue further to see how variables changed in relation to one another and whether there were any obvious gaps in this picture.

In this chapter, I describe the results of phase I and II of the project, corresponding to results of the screening and categorization phases, respectively. Here I describe both the results of this systematic search and comment on observations that I had made during the recording of these variables.

3.2 Methods

I would like to acknowledge the following people for helping with screening of studies: Cristina Nunes-Fonseca (CF) for being my second screener and Hanna Vesterinen (HV) for third screening any disagreements. I would also like to acknowledge Angus Sinclair (AS) and Alexandra Bannach-Brown (ABB) for their help with categorization of a subset of the studies included.

Methods are described in detail in the previous chapter, Chapter 2. Results are described here primarily from Phase I of the project, which involved performing an electronic search of the literature and having two independent investigators screen these identified studies for relevance based on their title and abstract. I also include data collected as part of Phase II of the project, where included studies from Phase I were screened for a second time at full-text level for categorisation according to methods used for model induction, outcomes measures within experimental designs and treatment administered, if any given.

3.3 Results

3.3.1 Phase I. Search of the literature

The search performed in PubMed identified 14,721 publications (Figure 3.13.2, all data is available through request at <https://doi.org/10.5281/zenodo.1209832>). Two independent investigators double screened all publications. Using phase I inclusion and exclusion criteria 3461 publications were agreed to have potential relevance by both screeners. A remaining 9625 studies were agreed to be excluded by both investigators at this stage of the project. 1436 studies had to be refereed by a third investigator due to disagreements between screener 1 and 2. Of this, 874 were concluded to be of relevance. 199 search results were excluded from any further stages of the project, as no abstracts could be obtained. 4335 publications advanced further into phase II of the project.

Categorization of studies at an abstract level identified 89% of studies included in this initial phase of the project reporting to be modelling schizophrenia in their animal experimental paradigms (Figure 3.23.1). Other publications referred to the modelling of substance-induced psychotic disorders or psychotic disorders due to a medical condition in the literature. 4% of publications did not specify a condition their study was of relevance to.

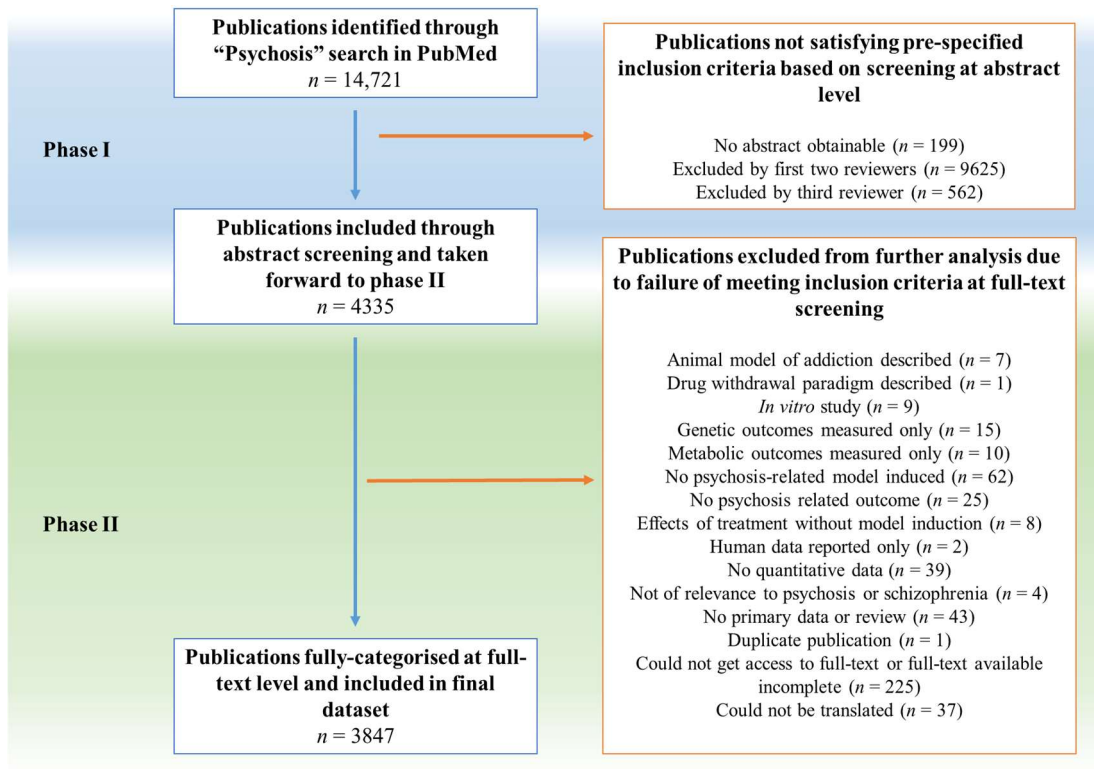


Figure 3.13.2 Flow diagram of publication inclusion

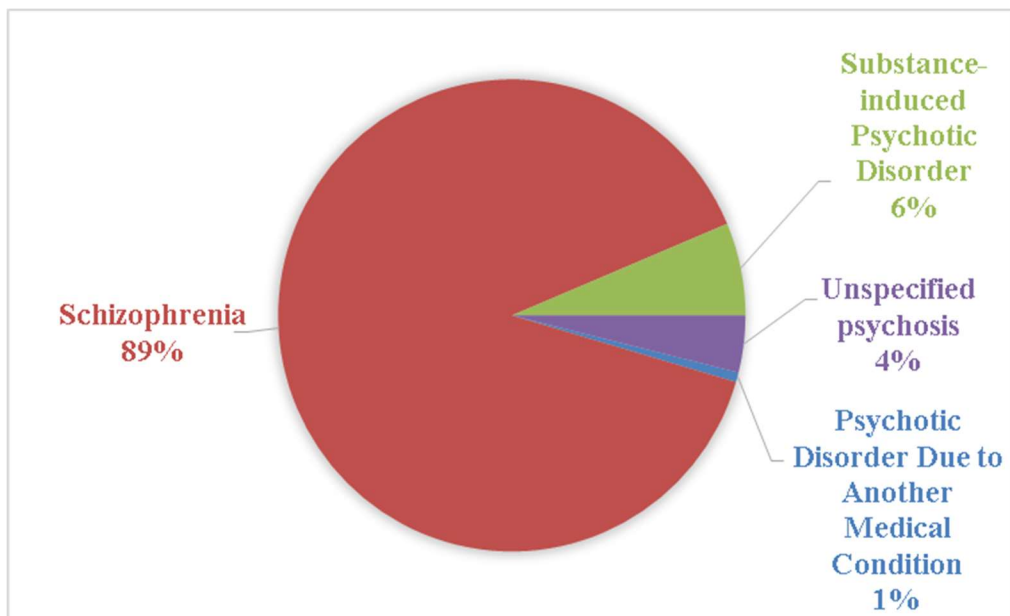


Figure 3.23.1 Phase I categorisation of studies at abstract level

3.3.2 Phase II. Categorisation of studies of relevance

3.3.2.1 Summary of experimental design after categorisation of literature

During phase II of the project, a further 451 publications were excluded for not meeting the inclusions criteria that were pre-specified. 37 publications were categorised as potentially being of relevance, however, in a foreign language, which could not be translated and therefore were excluded from further stages of the project. 3847 studies were included in phase II of the project and fully categorised. Categorisation of each of these studies included recording information about experimental approaches used to establish the model of interest, outcomes measured and used to assess animal models with, and treatment, if any, administered in an effort to alleviate the model. I describe these in more detail below. In addition, studies were also categorised according to the reporting of a list of criteria describing risk of bias and other methodological quality items. The results of this categorization are presented in Chapter 6.

3.3.2.1.1 Experimental approaches to modelling

Experiments describing animal models of psychotic disorders were broadly clustered into the four different groups mentioned in the literature: pharmacological-, genetic-, developmental- and lesion-induced models (Jones et al., 2011). Studies reported using one or a combination of these methods for model induction. The lesion group encapsulated models using high frequency stimulation, temporary inactivation and full lesions. For methods used, which did not quite fit into any of these four categories, I created additional categories. I created a category labelled „Environmental” to describe interventions that I considered environmental triggers in humans, as opposed to pathologic risk factors. This category mainly involved various forms of inducing stress in mature animals. As predominantly it is thought that schizophrenia is an early neurodevelopmental disorder (Corcoran et al., 2002), I considered environmental influences in later life as being different to those applied during early development of the animal. Examples of this latter method were categorised as developmental animal models instead. I only considered ‘environmental’ inductions to be a valid animal model if they had been combined with another method of model induction. Similar to this, ‘Adjunctive’ models included interventions that by themselves would not be sufficient to induce an animal model

of psychosis, however, when given supplementary to another intervention such as a pharmacological intervention, it could be considered a valid model.

Further additional categories were modelled on classifications presented in ICD-10 and DSM-5, namely substance-induced psychotic disorder, medication-induced psychotic disorder and psychotic disorder due to another medical condition. Those initially categorised as animal models of substance-induced psychotic disorder were eventually combined with other pharmacological models as many pharmacological models of schizophrenia are also considered to be substances of abuse in humans (Steeds et al., 2015). The final two categories were labelled as 'puerperal psychosis' and 'menstrual psychosis', neither of which are officially recognised by current psychiatric nosology, however, these were described in the literature as separate entities and therefore categorised here as such. Of course, publications often reported on more than one of these models within the same study, and therefore all possible reports of a model were recorded. 762 studies (20% of all included studies) reported combining some of these models to measure their effects in animals, and these I regarded as combination models in later stages of the project.

Through this broad grouping of induction methods it is very clear that the models most often reported in the literature were pharmacologically induced (Table 3.1). The most common method of inducing a model was through the administration of a pharmacological agent. There were 4517 reports of pharmacological intervention

Table 3.1 Prevalence of model induction method used to induce animal model of psychosis reported in the literature

Induction category	Number of times reported in the literature	Percentage of all models
Pharmacological	4517	67.4%
Developmental	877	13.1%
Genetic	838	12.5%
Lesion	211	3.1%
Environmental	137	2.0%
Observational or Trained	68	1.0%
Psychotic Disorder Due to Another Medical Condition	29	0.4%
Menopausal psychosis	16	0.2%
Adjunctive	6	0.1%
Medication-induced Psychotic Disorder	3	0.04%
Puerperal psychosis	1	0.01%

used in the literature, accounting for 67.4% of all methods reported in the literature. Lesion studies were the least widely reported animal models of schizophrenia in the literature out of the original main four categories for models of schizophrenia.

I wondered if these observations might be explained by these models having been used for longer than other model induction methods. When looking at changes in reporting over time, data indicates that pharmacological models have predominated the pre-clinical research field of psychotic disorders (Figure 3.3). While in recent years the use of genetic and developmental models has increased, pharmacological manipulations have continued to be at the forefront of research ahead of all of the other models reported in the literature.

Overall, through categorisation of the literature I found 852 different ways reported to induce or potentiate animal models of psychotic disorders (see *Appendix II. Results of categorisation of the literature: Complete list of model induction methods used, behavioural outcomes measured, and treatment compounds tested*). Pharmacological models accounted for the largest proportion of these models (Table 3.2). In total 309 different pharmacological models were recorded to be reported in the literature. This was followed by 292 different genetic interventions and 129 different ways of inducing developmental animal models of schizophrenia. Many of the individual methods used to create these models were reported in the literature only a handful of times.

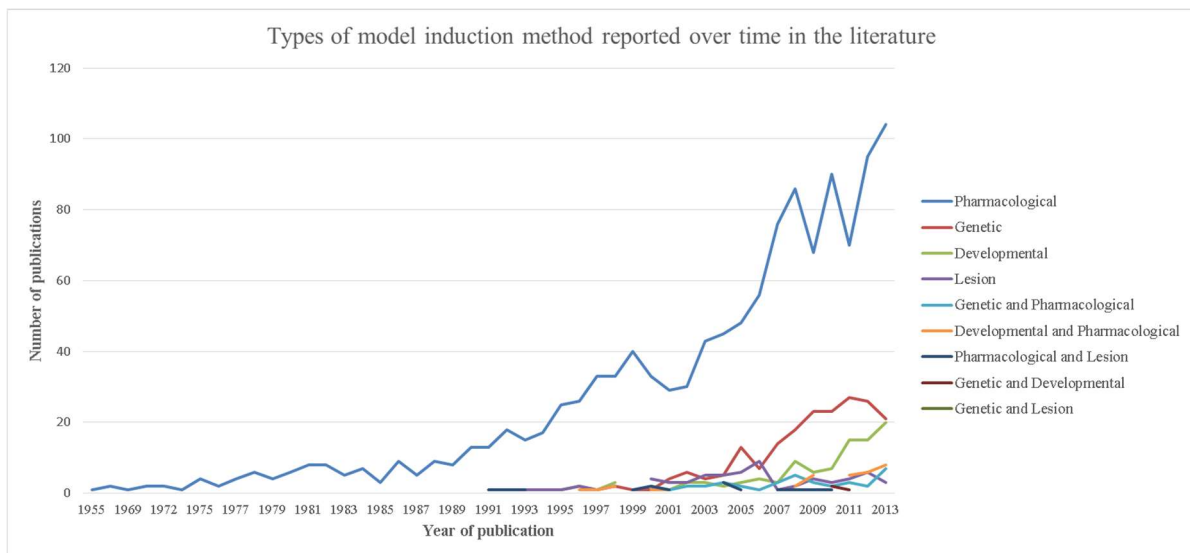


Figure 3.3 Prevalence of types of model induction methods reported in the literature over time

Table 3.2 Methods used to induce model of psychosis categorised

Induction category	Number of different models reported in the literature	Percentage of all models
Pharmacological	309	36.3%
Genetic	292	34.3%
Developmental	129	15.1%
Lesion	73	8.6%
Environmental	20	2.3%
Psychotic Disorder Due to Another Medical Condition	19	2.2%
Observational or Trained	3	0.4%
Adjunctive	3	0.4%
Medication-induced Psychotic Disorder	2	0.2%
Menopausal psychosis	1	0.1%
Puerperal psychosis	1	0.1%

As mentioned, many pharmacologically induced animal models of psychosis used substances of abuse and therefore could also be classed as substance-induced psychotic disorder (Figure 3.4).

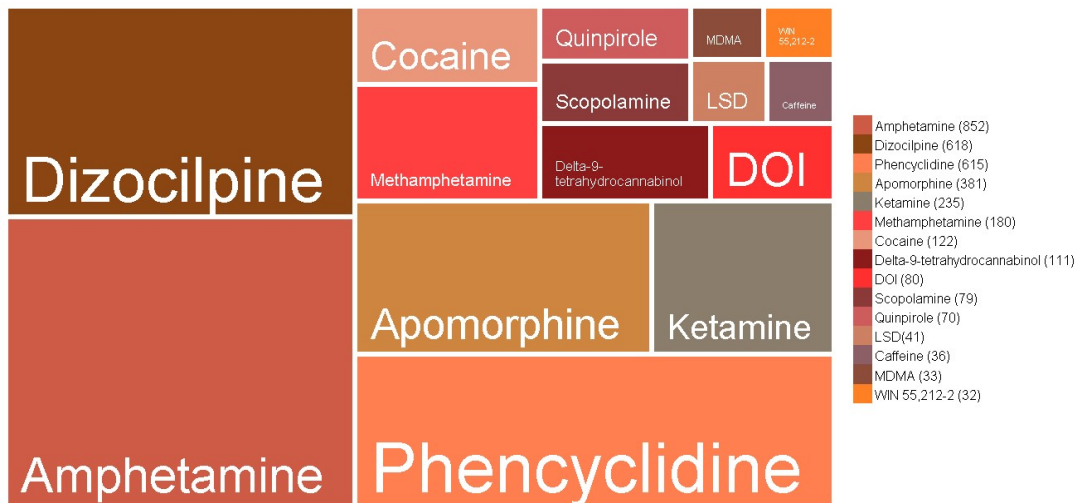


Figure 3.4 Top 15 interventions used to induce an animal model of psychosis pharmacologically

Numbers in brackets show the number of publications reporting these models in the collated literature.

Genetic models were the second most widely reported group of methods used to induce animal models of psychosis in the literature. The number of different genetic manipulations to create these animal models was high and accounted for about a third of all models in the final list (Figure 3.5).



Figure 3.5 Top 15 genetic manipulations used to induce an animal model of psychosis

Numbers in brackets show the number of publications reporting these models in the collated literature.

Developmental models were almost as widely reported as genetic models in the literature, however, the number of actual models within this group was much less. This group mainly involved pre- and postnatal infections of pups as well other early disruptions to development, including obstetric complications and early damage to the brain (Figure 3.6).

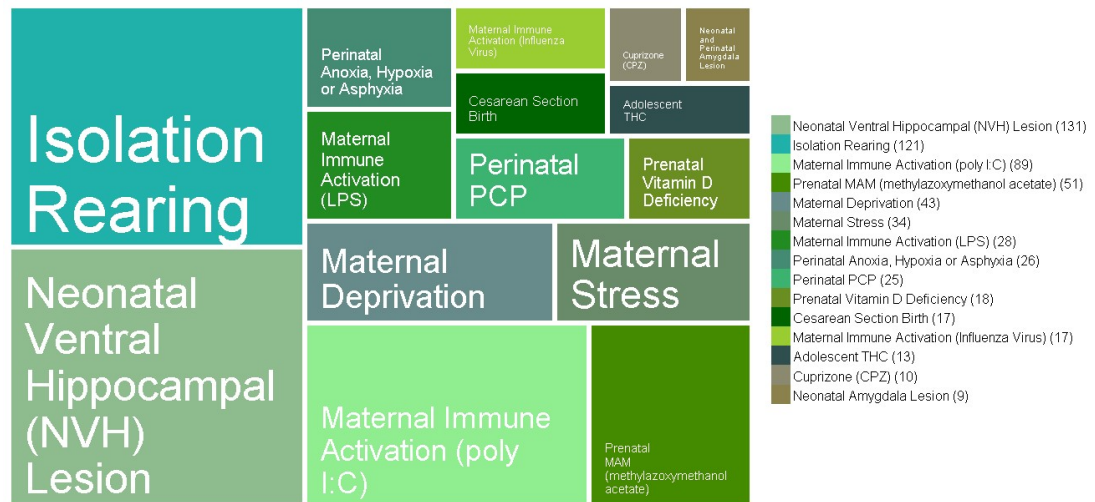


Figure 3.6 Top 15 interventions used to induce a developmental animal model of psychosis

Numbers in brackets show the number of publications reporting these models in the collated literature.

Lesion models were far less widely reported in the literature compared to the other main groups of induction methods. Lesions in this category were mostly introduced in adulthood, as any lesions given at a young age were categorised as developmental animal models.

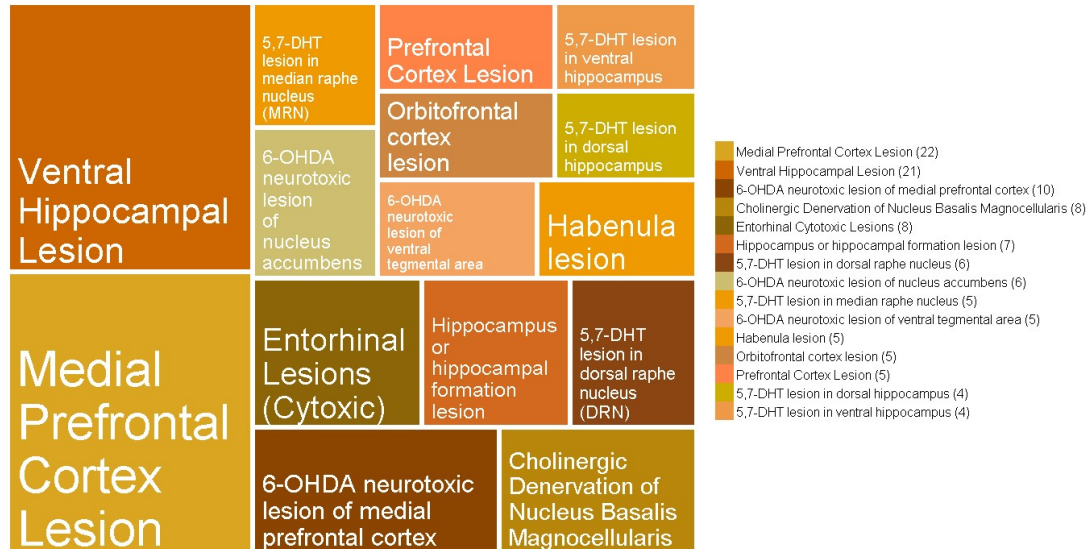


Figure 3.7 15 most widely reported lesion models in the preclinical psychosis field

Numbers in brackets show the number of publications reporting these models in the collated literature.

3.3.3 Outcome measures reported in the literature

Categorising studies according to type of outcome they were measuring revealed that most studies reported behavioural outcomes. In total, 2951 studies (77%) reported measuring effects on behaviour, 358 studies (9%) reported anatomical outcomes, 476 studies (12%) reported electrophysiological and 1601 studies (42%) reported neurochemical outcomes.

Behavioural measurements were further classed according to the test that was being used, the corresponding behaviour this test was thought to measure in the animals and the human behaviours that these animal behaviours are thought to have relevance to.

The most common human symptoms to be modelled in animals were psychomotor agitation, anxiety and sensorimotor gating. Psychomotor agitation was based on measurements of animal behaviours such as locomotor performance and stereotyped behaviour and was reported in 2488 publications (84% of all studies) (Table 3.3). Other widely reported measures included measures relevant to negative symptoms such as anxiety-like behaviour in animals, and behaviours of relevance to cognitive deficits such as learning and memory and, in some ways, sensorimotor gating.

Table 3.3 Human behaviours of relevance to outcomes reported to be measured in the preclinical literature

Category of Human Behaviour thought to be measured	Animal behaviours measured within this category	Number of publications reporting measurement
Psychomotor agitation	Motor performance, Stereotyped behaviour	2488
Anxiety	Anxiety-like behaviour	1455
Sensorimotor gating	Latent inhibition, Sensory gating	1227
Learning and Memory	Affective learning, Associative learning, Attention and Memory, Avoidance Learning, Discrimination learning and memory, Latent learning, Long-term memory, Reference Memory, Working memory, State-dependent retention, Relational memory	1165
Social behaviour	Social behaviour	451
Depression	Behavioural despair	148
Attention	Attention	120
Executive functioning	Cognitive flexibility, Problem-solving	91
Motor Co-ordination	Motor co-ordination	88
Pain sensitivity	Nociception	77
Avolition	Motivation, Reward-seeking behaviour	59
Anhedonia	Hedonic reaction to reward, Reward sensitivity	35
Impulsivity	Avoidance behaviour, Decision-making, Impulsivity, Risk taking	33
Hallucination	Hallucinatory-like behaviour	32
Other general abnormal behaviours	Natural behaviours	27
Aggression	Aggression	19
Olfactory dysfunction	Odour discrimination, Olfaction	17
Brain lateralization	Functional brain asymmetry	16
Psychotic polydipsia	Polydipsia	15
Empathy (evolutionary precursor)	D2 receptor priming	11
Interval timing	Timing	11
Tardive dyskinesia	Spontaneous orofacial movements	10
Insomnia	Sleep-wake pattern	7
Mother-infant interaction	Mother-infant interaction	6
Central nervous system and dopaminergic activity	Central nervous system and dopaminergic activity	3
Communication deficits	Communication	3
Epileptic-like outcome	Epileptic-like outcome	3
Human laughter	Positive affect	3
Discriminative stimulus properties of drugs	Discriminative stimulus properties of drugs	2
Sensitivity to cannabinoids	Sensitivity to cannabinoids	2
Taste sensitivity	Taste sensitivity	2
Activation of the opioid receptor system	Activation of the opioid receptor system	1
Catatonia	Catatonia	1
Perseveration	Perseverative behavior	1

Furthermore, data shows that this measure of behaviour has continuously dominated the pre-clinical research field among all other measures of behaviour in animal models of psychotic disorder over the years (Figure 3.8).

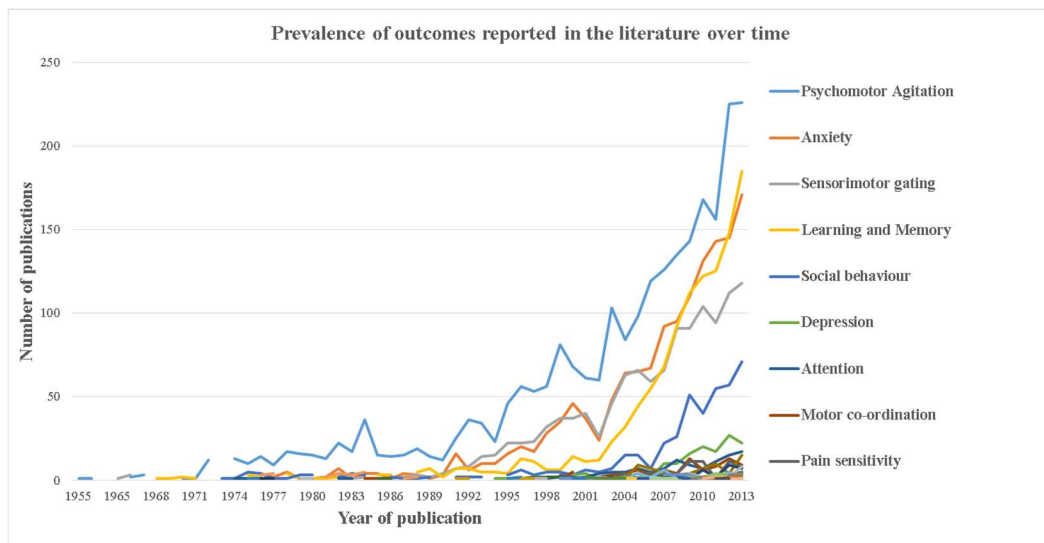


Figure 3.8 Prevalence of human behaviours of relevance to outcomes being assessed in animal models of psychosis over time

In total 336 different behavioural outcome tests were recorded to be reported in the literature. **Figure 3.9** shows the 10 most commonly reported tests in the literature to measure psychomotor agitation (for the full list of behaviours, see *Appendix II. Results of categorisation of the literature: Complete list of model induction methods used, behavioural outcomes measured, and treatment compounds tested*). These mainly included a measure of locomotor activity and other measures of stereotyped behaviour.

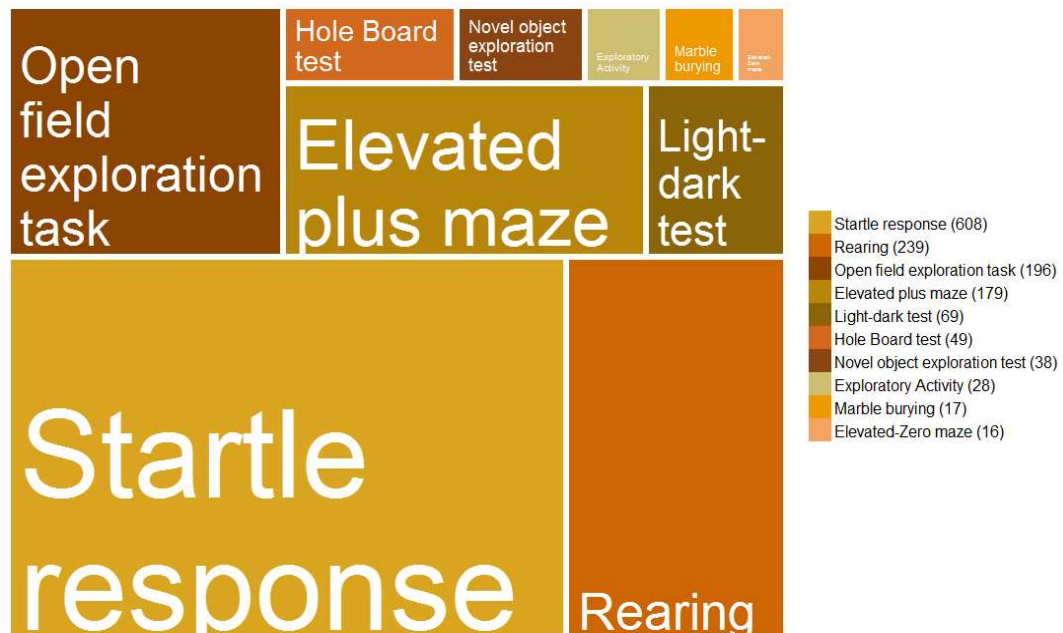


Figure 3.9 10 most commonly reported measures of animal behaviour of relevance to psychomotor agitation

Numbers in brackets show the number of publications reporting these measures in the collated literature.

Anxiety-like behaviour was the second most widely reported measure in the literature.

Figure 3.10 shows the 10 most commonly reported tests in the literature to measure



this form of behaviour.

Figure 3.10 10 most commonly reported anxiety-like behaviours in the preclinical literature

Numbers in brackets show the number of publications reporting these measures in the collated literature.

Deficits in sensorimotor gating was mainly measured through pre-pulse inhibition, but also some other tests. Figure 3.11 shows the 10 most commonly reported tests in the literature to measure this form of behaviour.



Figure 3.11 10 most commonly reported behavioural measures in animal models of psychosis of relevance to sensorimotor gating

Numbers in brackets show the number of publications reporting these measures in the collated literature.

Finally, disrupted learning and memory in animal models of psychosis was measured using a wide variety of different tests. **Figure 3.12** shows the 10 most commonly reported tests in the literature to measure this form of behaviour. Many of these tests were described in different ways, however, could have been otherwise different versions of the same test.

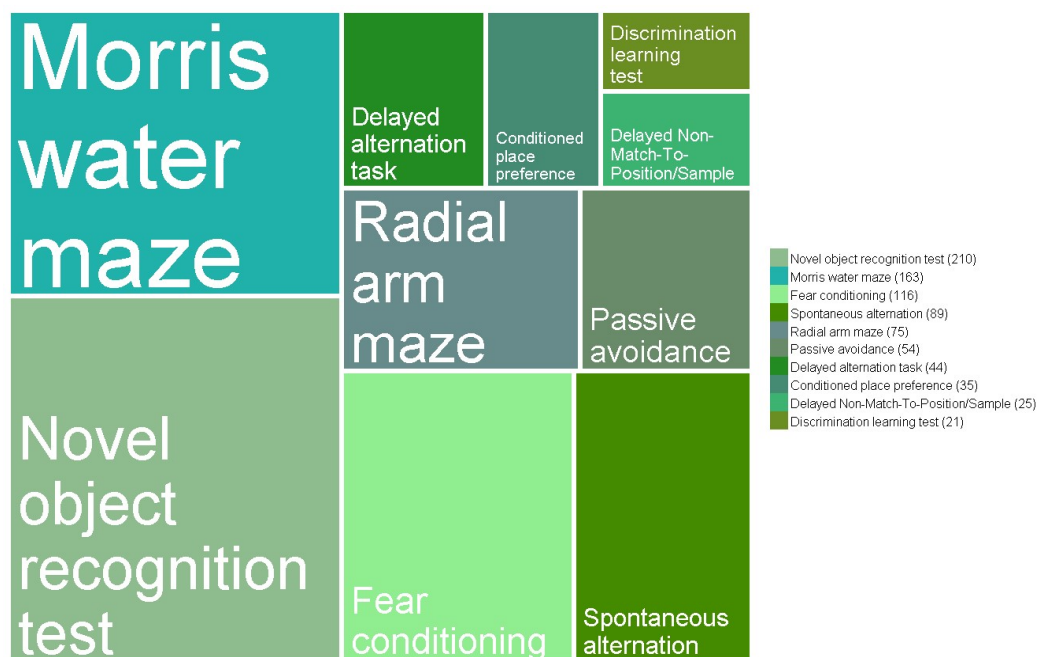


Figure 3.12 10 most commonly reported behavioural measurements in animal models of psychosis of relevance to deficits in learning and memory
Numbers in brackets show the number of publications reporting these measures in the collated literature.

3.3.4 Treatments reported in the literature

I also included experiments in my review that looked at the effects of treatment drugs in these animal models. In total 1796 studies (47% of all studies) were categorized to have reported testing the effects of a drug in an animal model of a psychotic disorder. This identified a total of 946 different compounds in the literature as having been tested in animal models of schizophrenia.

The most commonly reported therapeutic agents in the literature are mainly understood to exert their effect on dopaminergic pathways in the brain and included many currently licensed antipsychotics (Table 3.4- showing top 25, for a full list of treatments see *Appendix II. Results of categorisation of the literature: Complete list of model induction methods used, behavioural outcomes measured, and treatment compounds teste*). Other pathways often targeted in the brain included modulations of neurotransmitters such as serotonin, glutamate, GABA, acetylcholine, and noradrenaline.

Table 3.4 25 most commonly reported treatments tested in animal models of psychosis in the literature

Drug	Mechanism of action	System affected	Number of publications
Haloperidol	Dopamine D2 receptor antagonist	Dopamine	586
Clozapine	Serotonin 5-HT _{2A/2C} receptor antagonist and dopamine D _{2/4} receptor antagonist	Dopamine and Serotonin	464
Risperidone	Dopamine D2 receptor antagonist and serotonin 5-HT _{2A} receptor antagonist	Dopamine and Serotonin	168
Olanzapine	Dopamine D _{2/3/4} receptor antagonist, serotonin 5-HT _{2A/5-HT2B/5-HT2C/5-HT6} receptor antagonist and α -1 adrenergic receptor antagonist	Dopamine, Noradrenaline and Serotonin	113
SCH23390	Dopamine D1 receptor antagonist	Dopamine	69
Chlorpromazine	Dopamine D2 receptor antagonist	Dopamine	64
Sulpiride	Dopamine D2 and D3 receptor antagonist	Dopamine	48
Raclopride	Dopamine D _{2/3} antagonist	Dopamine	44
MDL 100907/M100907	Serotonin 5-HT _{2A} receptor antagonist	Serotonin	44
Nicotine	Nicotinic acetylcholine receptor agonist	Cholinergic	42
Aripiprazole	Serotonin 5-HT _{1A} receptor agonist, serotonin 5-HT _{2A} receptor antagonist and dopamine D _{2/3} receptor agonist	Dopamine and Serotonin	42
Quetiapine	Serotonin 5-HT _{1A} /5-HT ₂ receptor antagonist, dopamine D _{1/2} receptor antagonist, also binds to other α -1, α -2 adrenergic and histamine H ₁ receptors	Dopamine and Serotonin	39
Diazepam	GABA A receptor antagonist	GABA	39
SR141716/Rimonabant	Cannabinoid CB ₁ receptor antagonist	Cannabinoid	30
LY379268	Glutamate mGluR 2/3 receptor agonist	Glutamate	27
D-Serine	Co-agonist of glutamate NMDA receptor	Glutamate	26
Ritanserin	Serotonin 5-HT ₂ receptor antagonist	Serotonin	25
Ketanserin	Serotonin 5-HT ₂ receptor antagonist	Serotonin	25
WAY 100635	Serotonin 5-HT _{1A} receptor antagonist	Serotonin	25
Prazosin	Adrenergic α -1 adrenoceptor antagonist	Noradrenaline	22
8-OH-DPAT	Serotonin 5-HT _{1A} receptor agonist	Serotonin	21
Ziprasidone	Dopamine D2 receptor agonist and serotonin 5-HT ₂ receptor antagonist	Dopamine and Serotonin	21
Fluoxetine	Selective serotonin reuptake inhibitor	Serotonin	21
LY354740	Glutamate mGlu2/3 receptor agonist	Glutamate	20

Drugs thought to modulate the dopamine system were most commonly reported in the literature. Figure 3.13 shows the 15 most commonly reported drugs acting on this neurotransmitter.

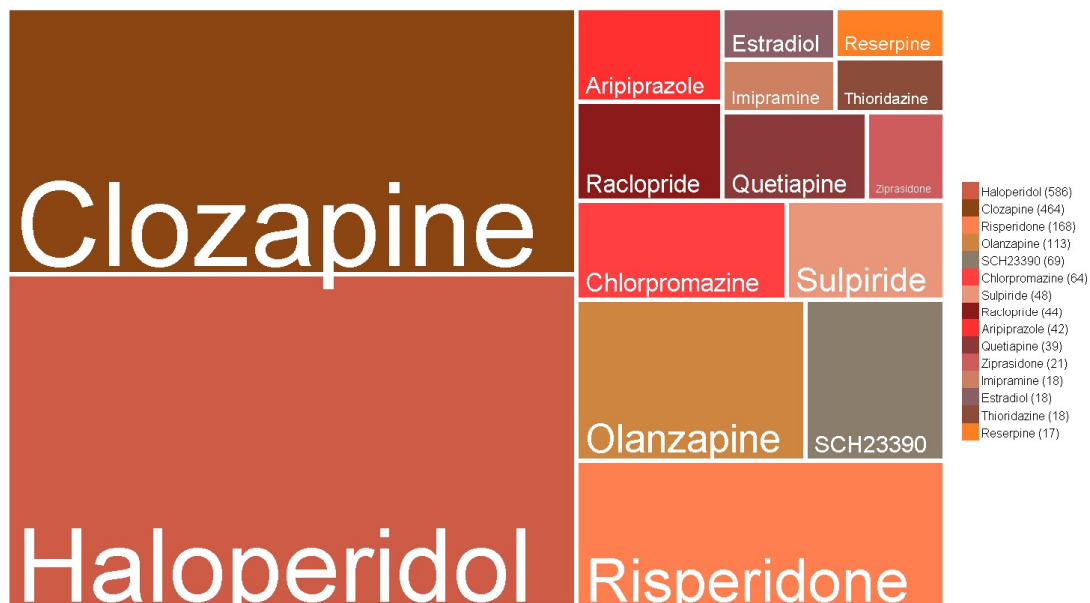


Figure 3.13 15 most widely reported dopaminergic treatment drugs in the literature Numbers in brackets show the number of publications reporting these treatments in the collated literature.

Modulation of serotonin was also popular in the field as a number of drugs administered as therapeutic compounds are thought to exert their action on serotonin receptors. Figure 3.14 shows the 15 most commonly reported drugs acting on this neurotransmitter.

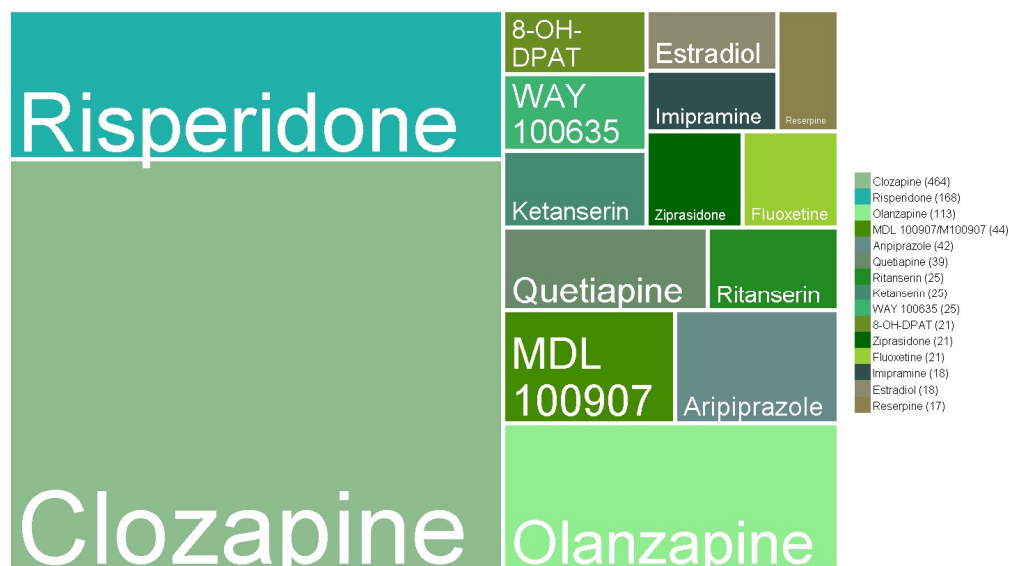


Figure 3.14 15 most widely reported serotonergic drugs treatment drugs in the literature
Numbers in brackets show the number of publications reporting these treatments in the collated literature

Finally, glutamate is another widely studied pathway in animal models of psychosis and Figure 3.15 shows the 15 most commonly reported drugs acting on this neurotransmitter.

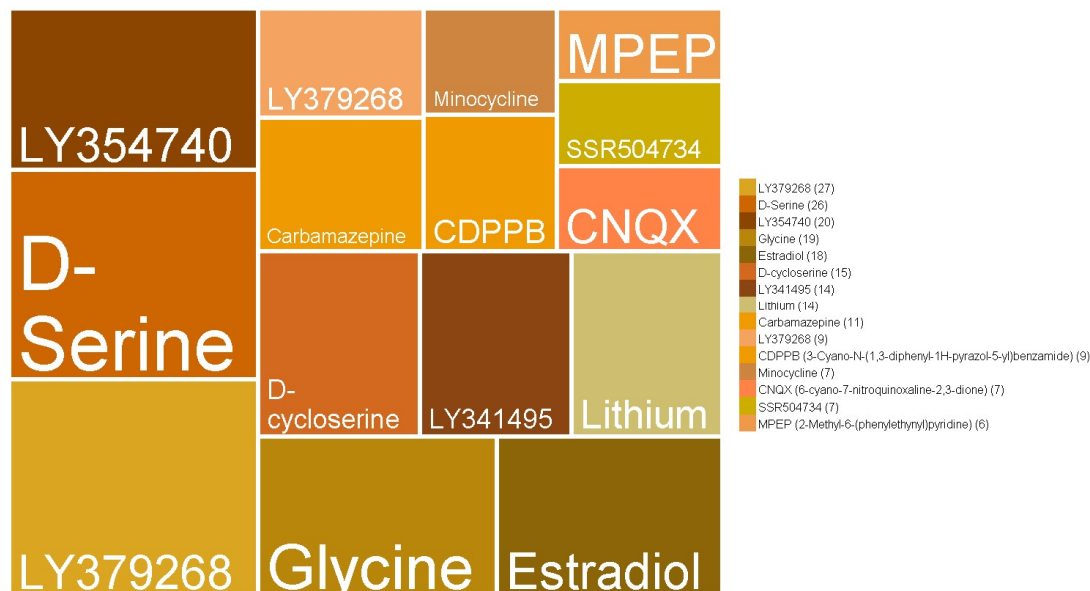


Figure 3.15 15 most widely reported glutamatergic treatment drugs in the literature
Numbers in brackets show the number of publications reporting these treatments in the collated literature.

In total 734 treatments (76%) were categorized as having been reported a twice or less, with 587 of these reported only once (62% of total treatments reported).

3.4 Discussion

A broad search of the literature identified a substantial amount of studies of potential relevance to animal models of psychosis and psychotic disorders. The 29% of publications included to be of relevance after the second phase of full-text screening was higher than initially expected, however, likely a result of the broad inclusion/exclusion criteria. On this initial screening of identified studies at an abstract level we categorised publications to be overwhelmingly reporting animal experiments of relevance to schizophrenia. Very few studies could be classed as anything other than this.

3.4.1 Methods used to induce the model

Through categorization of publications in phase II of the project, over 800 different inductions were recorded to be reported as of relevance to psychosis or more specifically schizophrenia. This number is volumes bigger than what is normally referenced in the literature in reviews of animal models of schizophrenia. A summary resource published by the Schizophrenia Research forum that aims to provide a comprehensive list of animal models used for research in schizophrenia, referenced 149 different models (Koenig, 2014).

This substantial difference might be explained by the approach of this review when categorizing models. As a broad review I wanted to capture the field as it is and portray it without any subjective selection bias on my part. As a result the compilation presented here is a list of all animal interventions that have been reported in the literature to bear relevance to a psychotic disorder (in most cases schizophrenia) in the clinic. Arguably perhaps some of the models identified here would not be classed as classical models of schizophrenia *per se*, but of course this interpretation is subjective and there are no pre-defined criteria on this (Nestler and Hyman, 2010). Of course animal modelling for psychiatric disorders is especially challenging because the boundaries between different disorders especially their clinical profiles are often hazy and subjective (Hyman, 2010). Moreover, due to our limited understanding of these disorders in terms of pathophysiology, a lack of valid biomarkers and objective diagnostic tests, it is difficult to create true models of these

human disorders. Nevertheless, in the literature it is a common acceptance that animal models of schizophrenia are unlikely to and therefore are not required to model the entire spectrum of symptoms associated with the disorder to be classed as a valid animal model of schizophrenia (Fernando and Robbins, 2011). As a result authors often refer to their animals as models of schizophrenia even when they are simply assessing for signs believed to be representative of the positive symptoms in schizophrenia. In fact it is recognised in the literature that an animal model of schizophrenia can be used for the purpose of further understanding a specific aspect of the disease, further exploring the validity of risk factors or for testing the efficacy of potential therapeutic drugs (Moore, 2010). For these reasons I did not exclude any models reported to be of relevance in the field. Unfortunately this relaxed approach to modelling also means that some animal models identified here might not be specific to schizophrenia and experimental setups might have relevance to other psychiatric disorders of similar etiology, genetic risk or clinical profiles (Doherty and Owen, 2014). This supports the belief held by some researchers in the field that a lack of progress in psychiatric research thus far can likely be attributed, at least in part, to current polythetic – categorical classification systems (Hengartner and Lehmann, 2017). In fact, it was often observed that a publication would report an experimental model setup as being of relevance to multiple psychiatric disorders, one of them being a psychotic one. How differential labelling of studies being of relevance to multiple disorders affects the field and conclusions drawn in reviews as this one, I discuss later in Chapter 7.

3.4.1.1 *Pharmacological models*

Overall, most models recorded could be classed as pharmacological interventions and these were also the most widely reported group of models in the literature. A lot of models of substance-induced psychosis fell into this category as many stimulant drugs are used to induce animal models of relevance to schizophrenia (Steeds et al., 2015). Data collected here shows that these models have predominated the preclinical research field since the 1950s. The popularity of pharmacological methods to induce animal models of schizophrenia stems from early observations by Young and Scoville in 1938 that psychostimulants like amphetamines can induce a psychotic-like disorder in humans (Potvin et al., 2005). These animal models are based on our understanding of how different neurotransmitter systems may be affected in schizophrenia and therefore are thought to have good construct validity

for the disorder (Marcotte et al., 2001), despite the fact our understanding of the underlying biology is still very much limited. Due to the schizophrenic-like symptoms these drugs are able to create or exacerbate when administered in humans, these models are thought to also have good predictive validity for schizophrenia (Marcotte et al., 2001). Although it is important to note that in the clinic in some cases psychotic symptoms only arise after chronic administration of these drugs (e.g. amphetamine) (Steeds et al., 2015), whereas animal models often employ acute experimental paradigms using the same drugs. One major limitation of these models is their lack of etiological validity – in that they are unable to induce changes across multiple different neural systems, which is thought to be part of the complex pathophysiology of the disorder (Nestler and Hyman, 2010). Importantly, any treatment compounds discovered through the reversal of the effects of these psychotic pharmacological agents are constrained by the pharmacology of these agents to the specific mechanisms they are acting on (Moore, 2010). This means that novel treatments found to work in these models are compounds that are simply able to attenuate or reverse the specific effects of the manipulation itself (Wilson et al., 2010). This has limited the utility of these models in finding treatments with novel routes of action and treatments for negative and cognitive deficits, and as a result, novel therapeutics developed in these models have led to a large number of false positive compounds that show little or no efficacy in clinical trials (Moore, 2010).

Another major limitation of these models is that they often fail to recapitulate the clinical nature of schizophrenia, which is a chronic, neurodevelopmental disorder, marked by transient combinations of different symptoms predominating at different stages (Steeds et al., 2015). Nevertheless, these models have been popular because they are easy and quick to perform and have been influential in forming the three most well-established theories of schizophrenia: the dopamine, the serotonin and the glutamate hypotheses (Geyer and Moghaddam, 2002).

3.4.1.2 Genetic models

As our understanding of the underlying biology of schizophrenia has increased over recent years in terms of what pathophysiological, genetic and environmental risk factors are associated with the disorder, models using lesion, genetic and developmental methods of 'schizophrenia' induction have also been introduced. We see from the data here that these models only really became more popular in the literature from 1990s onwards. Genetic and developmental models are mostly

examples of “risk factor models” whereby methods used to induce the model are based on our knowledge of susceptibility genes for schizophrenia or early environmental risk factors such as increasing paternal age, prenatal immune activation or stress of the mothers and other adverse early life experiences (Moore, 2010).

Genetic models have included complete knockouts, heterozygotes or conditional knockouts of genes of relevance, as well as transgenic overexpression of some genes, which tend to be based on post-mortem observations of genes upregulated in schizophrenic individuals. Limitations of these approaches are that the functional effects of genetic variants associated with schizophrenia are not always well understood. The association between the gene manipulated in a model and the risk for schizophrenia is not robust and therefore a knockdown or knockout of the gene in an animal is not necessarily homologous to the variation seen in humans with schizophrenia (Moore, 2010; Powell and Miyakawa, 2006). A recent GWAS study of schizophrenia did not find any common genetic variants that contributed a meaningful effect at genome-wide levels of significance (Ripke et al., 2013). While about 8,300 independent SNPs have been estimated to contribute to the aetiology of schizophrenia collectively through the same study, individually these alleles have a weak effect on schizophrenia risk. Exome sequencing studies have also not found significant support for any individual genes (Purcell et al., 2014) and early discoveries of candidate genes related to schizophrenia have also not been supported by these larger studies (Nutt and Need, 2014). Any possible candidate genes confirmed to be of relevance in future, will likely only be relevant to a small proportion of individuals with schizophrenia (Nutt and Need, 2014). As schizophrenia is thought to involve a large number of different genes, models focusing on single genes are unlikely to faithfully model the disorder in the laboratory. Nevertheless, genetic models can still be informative, but arguably only to a certain extent.

3.4.1.3 Developmental models

Developmental models were equally as prevalent in the literature as genetic models. Given the limited translational relevance of animal models of schizophrenia using pharmacological or genetic interventions, and the amount of evidence supporting the neurodevelopmental hypothesis, developmentally impaired animal models seem like a promising approach to improving understanding of underlying pathophysiology

and translating these insights into the clinic. These models are based on evidence that suggests that adverse environmental hits, especially during early life increases the risk of the development of schizophrenia (Dean and Murray, 2005). The observations that symptoms usually arise in adolescence or early adulthood further support the belief that the underlying pathology begins in early brain development (Powell, 2010). Advantages of these models include the absence of any confounding drug or surgical interventions when testing novel compounds of therapeutic value, so that therapeutic compounds, which act of multiple pharmacological mechanisms can still be detected. This would allow for the discovery of treatment drugs which work differently to established medication and would potentially treat symptoms not being managed adequately by current medication (Jones et al., 2011). Moreover, developmental models offer a unique condition for the investigation of schizophrenia during early life and the prodromal phase of the disorder before symptoms arise, potentially leading to the identification of disease-modifying agents (Geyer et al., 2012). One major downside to these models is their lack of specificity to schizophrenia or any other psychiatric disorder for that matter (Nestler and Hyman, 2010). For example, early life stress is a risk factor for a multitude of psychiatric conditions in later life (Carr et al., 2013). As previously mentioned, where stress was not administered in early life I classed these interventions under the “environmental models” category. I believe these disturbances do not necessarily cause the disorder and in many cases do not lead to the manifestation of psychotic symptoms, however, they can trigger existing vulnerabilities and so if coupled with these then they too can contribute to the manifestation of the disorder (Corcoran et al., 2002). The face, construct and predictive validity of three commonly used developmental animal models are explored in further detail in Chapter 4.

3.4.1.4 Lesion models

The fourth major group of animal models of schizophrenia are lesion models, which have been far less frequently reported in the literature compared to the other models mentioned above. Lesion models are usually created to try to recapitulate the neurodevelopmental and neurodegenerative theories of schizophrenia and involve targeted lesioning of brain tissue in animals using electrolytic, aspiration lesions or chemically induced lesions through the use of excitotoxic agents (Marcotte et al., 2001). Research implicating the prefrontal cortex, hippocampus and thalamus in

schizophrenia has meant that many lesion models have focused on these areas of the brain in preclinical schizophrenia research (Marcotte et al., 2001). In this category I mainly classed lesions induced in adulthood, as I believe any lesions in the brain created during early development is likely to affect early maturation of the brain and should therefore instead be classed as a developmental model. This includes the widely reported animal model using neonatal lesioning of the ventral hippocampus. This is not always categorised the same in the literature (Jones et al., 2011). An important distinction between lesions created during early development and in adulthood is that early lesions like other developmental models of schizophrenia are able to show a delayed onset of behaviours thought to be of relevance to human symptoms (Marcotte et al., 2001). Lesion models are thought to be of limited relevance to schizophrenia. Primarily because while anatomical abnormalities are observed in the brains of schizophrenic individuals (Karlsgodt et al., 2010), there is no evidence to suggest that these cause the psychotic disorder in question (Nestler and Hyman, 2010). Moreover these models show much more extensive damage than what is normally seen in human brains affected by schizophrenia (Marcotte et al., 2001).

3.4.1.5 Combination models

Administration of psychotomimetic agents is often combined with other models to create a multi-hit model thought to have more construct validity for complex psychotic disorders such as schizophrenia (Mattei et al., 2015). In the literature a mere 20% of publications reported combination models, however, showing that future studies could focus more on these experimental paradigms. There is a lot of discussion in the field about whether we need to move away from using simple models to study schizophrenia in animals. This has seen the increase in the use of combination models that combine more than one method of induction to try to recapitulate the complexity of the neurodegenerative schizophrenia phenotype that arises from an interplay of multiple altered genes, abnormal neurotransmission systems and environmental factors (Sarnyai et al., 2015). To account for these factors and the late adolescent onset of schizophrenia, a “multi-hit” model of schizophrenia has been proposed, where there is an early disruption to the development of the central nervous system, which produces a long-term vulnerability to other hits that will cause the manifestation of symptoms associated with the disorder. It is thought that the development of schizophrenia is a result of

the convergence and interaction of genetic, environmental and other vulnerability risk factors in a cumulative manner during critical periods of neurodevelopment (Davis et al., 2016). Therefore, these kinds of combination models are thought to be a reliable way of integrating these developmental, genetic and environmental factors that are thought to cause the pathogenesis of schizophrenia in humans and be able to produce more valid animal models of the disorder (Maynard et al., 2001; Sarnyai et al., 2015). These models are also beneficial in broadening our research targets to include more opportunities to study other symptoms of schizophrenia, potentially improving the clinical relevance of data collected in animal studies. For example, it is said that the cognitive deficits of schizophrenia precede symptoms of psychosis in many cases, by an average of 9 years (van Oel et al., 2002), and their treatment is usually associated with a better therapeutic outcome (Mintz and Kopelowicz, 2007). Despite these arguments for the possible superiority of combination models, simple models are still believed to have utility in research without fully recapitulating the disease (Pratt et al., 2012). Importantly, however, the predictive value of any of our current models is still not clear as they have not yet led to the development of any clinically approved therapeutics (Geyer et al., 2012).

3.4.2 Outcome measures reported

As with models, I attempted to capture the field as it was reported. While outcomes have been categorised according to what is commonly claimed in the literature in terms of what a behavioural outcome measure in these animals is measuring, this can vary from study to study. I also attempted to group behavioural tasks and measures according to what they were suggested to measure. Often it was very difficult to class some of these behavioural outcome measures because many studies would report the apparatus that they had used to measure behaviour with and not the actual behaviour that was being analysed (e.g. T-maze as opposed to a measure of spontaneous alternation). Moreover, I found that many studies would report the same outcome measures using different names and experimental setups, or reporting it as measurements of different concepts making comparisons between different experiments and drawing conclusions from a group of similar experiments difficult. It is imperative that future studies make it clear exactly what aspect of a disorder is being studied so that the results and conclusions drawn have maximum predictive validity.

3.4.2.1 Measures of relevance to positive symptoms

Positive symptoms, which include hallucinations and delusions in humans, are difficult to recapitulate in animals as in humans they are diagnosed through verbal reporting and therefore it is not clear how effectively animal model-based biomarkers are able to model these symptoms (Steeds et al., 2015). Some studies refer to hallucinatory-like behaviour, however, of course this is a highly subjective measure of behaviour. Other animal behaviours thought to be analogous to positive symptoms include a measure of locomotion (i.e. usually a measure of hyperactivity) and stereotyped behaviour. While this behaviour does not correspond to any cardinal symptoms of schizophrenia (Nestler and Hyman, 2010) it is thought to be of relevance to psychomotor agitation in human schizophrenic individuals (Powell and Miyakawa, 2006). Changes in this domain of behaviour in animals is often measured in response to psychostimulants, NMDA receptor antagonists or novelty. Data collected here shows that this is the most widely reported measure of behaviour in the preclinical field of psychotic disorders. The advantages of this measure of behaviour are that it is easily carried out and shows sensitivity to antipsychotic drugs. Therefore, while this measure of behaviour is thought to have good predictive validity for antipsychotic efficacy in humans, it also has a number of limitations. First, in the clinic psychomotor agitation in schizophrenia is not overly common in individuals with only about 20% of all those experiencing schizophrenia presenting with episodes of agitation during their lifetime (Garriga et al., 2016). Moreover, locomotion is not a behavioural measure which is specific to the schizophrenia research field and is also widely used in other preclinical fields of research (Bailey and Crawley, 2009; Mchedlidze et al., 2011; Tatem et al., 2014). Moreover, as enhanced dopaminergic activity is thought to underlie the behaviour (van den Buuse, 2010), any treatment drugs that show response using this measure of behaviour are likely to only lead to the identification and development of more dopaminergic drugs.

3.4.2.2 Measures of relevance to negative symptoms

Negative symptoms are deficits in normal function and include symptoms of blunted affect, asocial behaviour, lack of motivation and impoverished speech (Nestler and Hyman, 2010). In the clinic negative symptoms are often diagnosed through interview-based approaches or through self-reporting (Barnes et al., 2014), which is

not something that is easily modelled in animal studies of course. Intrinsically human symptoms such as poverty of speech and blunted affect are especially difficult to model in animals (Sahin et al., 2016). However, research looking at ultrasonic vocalizations as shown here and discussed elsewhere, might be an experimental procedure in animals of some relevance to even unique behaviours like communication (Ferhat, Torquet, Le Sourd, 2016). Data collected here also shows that social behaviour and anxiety-like behaviour is widely reported in the literature, however, measures of affective state are far less prevalent. Measures of affective state and emotional regulation are important in human social interaction and communication (Sahin et al., 2016), therefore the strength of the inferences that can be made from animal social behaviour to human symptoms of social withdrawal are questionable. Current antipsychotics do little for improving negative symptoms associated with psychotic disorders, and development of new drugs is limited by our poor understanding of the underlying mechanisms mediating many of the symptoms in this domain (Young and Markou, 2015). Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia (CNTRICS) and Cambridge Neuropsychological Test Automated Battery (CANTAB) are two initiatives, which have been developed in recent years to aid the development of cross-species tests that can be used to study specific domains of schizophrenia symptoms (See www.cntrics.ucdavis.edu and www.cambridgecognition.com/cantab). So far, it has been recognised that studies trying to understand the neural mechanisms underlying behaviours of relevance to negative symptoms in animals have mainly used non-operant tasks (Young and Markou, 2015). It is believed that these neuromechanisms might be better understood by using more objective behavioural tests in preclinical studies that are similar to behavioural tests used in humans (Young and Markou, 2015). Finally, behavioural tests for negative symptoms are limited by the knowledge that specific negative symptoms may develop through varied mechanisms as they are not specific to schizophrenia and are implicated in other disorders like depression, autism and anxiety (Nestler and Hyman, 2010).

3.4.2.3 Measures of relevance to cognitive deficits

Cognitive deficits are described as core disturbances in psychotic disorders (Barch and Sheffield, 2014), and are associated with many other psychiatric disorders (Millan et al., 2012). They include impairments in attention and vigilance, working memory, reasoning and problem solving, processing speed, visual and verbal

learning and memory and social cognition (Nuechterlein et al., 2004). This domain of symptoms is also an important determinant of functional impairment and quality of life (Savilla et al., 2008). They precede all other symptoms and are relatively unresponsive to current antipsychotics available (Davidson et al., 2009). In recognition of this unmet clinical need and the need for the development of novel therapeutic compounds which address these symptoms in the clinic, initiatives such as the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS), CNTRICS and Treatment Units for Research on Neurocognition and Schizophrenia (TURN) programs were set up. These initiatives aim to identify important domains of cognitive impairment in schizophrenia and how to best measure and thus treat these in the clinic (Pratt *et al.*, 2012 and see www.MATRICS.ucla.edu). In animals it is extremely difficult to assess concepts such as thought or verbal learning and memory (Powell and Miyakawa, 2006) and therefore preclinical studies index certain cognitive function-related behaviours rather than be able to directly quantify it (Jones et al., 2011). There are no clear outlines for preclinical studies on cognitive test batteries, and therefore preclinical studies can often use behavioural paradigms, which have limited ability to measure the cognitive domains of interest accurately (Young et al., 2009). Animal behaviours thought to be of relevance to these symptoms collected from the literature here include various measures of learning and memory, which is the fourth most widely reported measure in the literature, along with measures of executive function and attention. While these are thought to have some extent of face validity for cognitive deficits seen in schizophrenia in humans, it is not always clear whether a cognitive task used in animals is measuring the same construct as is affected in humans (Pratt et al., 2012). Another major limitation of measures of cognitive deficits in animal models of psychotic disorders is that as with other measures of behaviours discussed above, they are not specific to schizophrenia. They are also seen in other neurological and psychiatric disorders; however, whether the disruptions in underlying neurobiological mechanisms are similar is not clear (Pratt et al., 2012). Cognitive tasks are also said to be confounded in animals by other factors such as low motivation or sedation reducing translational relevance of any results seen in these tests (Pratt et al., 2012). Developing and using tests, which are directly analogous to tasks that are used to measure the same constructs in neuropsychiatric test batteries in humans might improve this problem (Powell and Miyakawa, 2006). For example, a task normally used in humans for measuring

impulse control, namely the stop signal task has been recreated for use in preclinical experiments (Eagle and Robbins, 2003). Vice versa examples of tasks usually measured in animals, such as the Morris Water maze used for the assessment of spatial working memory has been modelled in a virtual version of this task in clinical research (Shipman and Astur, 2008). More similarities between the two research domains are likely to improve translation of knowledge from both sides.

3.4.2.4 Other measures of relevance

Additional abnormalities common in individuals diagnosed with schizophrenia such as deficits in sensorimotor gating have also been identified. A deficit in an operational measure of sensorimotor gating, namely pre-pulse inhibition (PPI), has been observed in many individuals in the clinic with schizophrenia (Braff et al., 2001).. Data collected here shows that sensorimotor gating is the third most widely described measure of behaviour in the preclinical literature. PPI has been proposed in recent years as a biomarker for schizophrenia (Mena et al., 2016). It is suggested that PPI is one of two neurophysiological measures, which fulfil all of the MATRICS/CNTRICS criteria and are suitable for use in clinical studies (Light and Swerdlow, 2014). This criterion expects that a neurophysiological biomarker is practical and stable over time, has utility as a repeated measure, is associated with functional outcome, has potential to show sensitivity to therapeutic agents, which is in line with observations in animals models and has clear links to neural circuits and behavioural mechanisms involved in the disease (Light and Swerdlow, 2014).

PPI is thought to be strongly driven by genetics in both animals and humans (Light and Swerdlow, 2014). Evidence shows that there is increased heritability in the decline of PPI measures in families with higher genetic vulnerability for schizophrenia (Greenwood et al., 2016). Relatives of individuals with schizophrenia and subjects with schizotypal personality disorder show deficits in PPI, similarly to schizophrenic individuals. These deficits are otherwise not seen in control subjects (Cadenhead, 2011; Cadenhead et al., 2000). Alongside recent studies in the literature suggesting that PPI levels in schizophrenia show long term stability (Light et al., 2012; Mena et al., 2016), this supports the idea that PPI deficits are a trait or a vulnerability marker of psychotic disorders. There is also some evidence to show that PPI levels correlate with deficits in neurocognition as well as global functional status (Swerdlow et al., 2006), and vary within individuals based on symptom state, whereby PPI levels improve with improvements in symptoms (Meincke et al., 2004;

Minassian et al., 2007). The literature also shows that PPI deficits are responsive to medication and are improved especially by atypical antipsychotics in schizophrenic individuals (Mackeprang et al., 2002; Oranje et al., 2002). This suggests that PPI deficits may also be a state marker of psychotic disorders. Therefore, disturbances in mechanisms that underlie PPI and information-processing, may both make an individual susceptible to developing psychosis and vary in relation to acute symptoms (Meincke et al., 2004).

This measure of behaviour has advantages over many other behavioural measures, as it is a phenomenon, which can be studied in animals and therefore holds good face and predictive validity for schizophrenia. It can be measured similarly in both rodents and humans, with similar results (Powell and Miyakawa, 2006), meaning this measure is able to increase our understanding of the neural and cellular substrates that underlie its translatability (Light and Swerdlow, 2014). Despite its suggested utility for predicting the likelihood of recovery in response to different types of therapies for cognitive interventions for schizophrenia (Light and Swerdlow, 2014), we must keep in mind its limitations. This measure of behaviour is not specific either to schizophrenia. It has been documented to occur in many other conditions including bipolar disorder, obsessive-compulsive disorder and Lewy body dementia (Geyer et al., 2001; Perriol et al., 2005; Perry et al., 2001), which limits its applicability as a biomarker specifically for schizophrenia or other psychotic disorders.

3.4.3 Therapeutic compounds reported to be tested

Overall, just under half of all studies identified to be of relevance in the literature reported studying potential therapeutic compounds in animal models of psychotic disorders. Overall, data collected here shows that the majority of treatments tested previously target the dopaminergic and serotonergic pathways. In fact, the top nine therapeutic compounds reported in the literature have all been established to have some affinity for dopamine receptors. Most of the drugs at the top of the list of most widely reported drugs are also currently established antipsychotics. Due to their serendipitous discovery, most targets of current anti-psychotic treatments and their mechanisms of action were discovered *post hoc* (Nestler and Hyman, 2010). Therefore older studies have used „back-translational psychopharmacology” in order to learn more about these drugs (Moore, 2010), while also some newer studies use these established drugs as reference treatments when testing the effect of novel

treatment options. This is based on the principle that we know these drugs work in humans therefore if we can find drugs that show similar effects in animals then they are likely to also do well in humans. Of course, this arguably only identifies more compounds that work in similar ways to these drugs and doesn't introduce drugs with novel mechanisms of action. Based on our knowledge of psychosis at this stage, we think that one out of the possible six dopamine pathways in the brain is affected in psychosis. Blocking of other dopaminergic pathways are what underlie other effects of dopaminergic drugs such as extrapyramidal side effects, secondary negative or cognitive deficits and hyperprolactinemia (Correll and Kane, 2014; Sesack and Carr, 2002). Unfortunately, a dopaminergic drug is unable to selectively target a single dopamine pathway and therefore targeting of other neurotransmitters are required to balance dopaminergic modulation in the other five dopaminergic pathways in the brain (Correll and Kane, 2014). Drugs of alternative mechanisms of action explored in the literature since these early drugs have included compounds acting on other pathways in the brain thought to be implicated in schizophrenia including cholinergic, glutamatergic, GABAergic, adrenergic, histaminergic or opioid pathways, as shown by data collated here. Many compounds interact with more than one of these pathways and therefore might be better candidates for treatment in the clinic as opposed to drugs, which target a single neurotransmitter system in the brain (Li et al., 2016).

Despite these alternate psychotropic drugs developed for treatment of psychotic disorders, the approval of these drugs has not been successful. Many drugs have shown efficacy in animal models, but have not gone on to show this same efficacy in humans (Moore, 2010). The United States Food and Drug Administration (FDA), only licensed drugs for a long time if they showed efficacy in positive symptoms of schizophrenia. This prevented the identification and development of drugs effective for other symptoms of the disorder such as negative symptoms and cognitive deficits as mentioned above (Geyer, 2006). So perhaps with increased focus from initiatives in recent years (Pratt et al., 2012) on the treatment of symptoms of psychotic disorders unmet by current medication and those which arguably affects quality of life the most in patients, novel drugs will be introduced. It has been highlighted that the main unmet needs in schizophrenia treatment currently include medications that can treat negative symptoms and improve cognition, can help treatment-resistant patients and those that will increase compliance (Fellner, 2017). It is thought that manipulation of the appropriate dopaminergic, serotonergic and

glutamatergic targets can help to manage a range of different symptoms including negative symptoms and cognitive deficits. Improving tolerability and safety of drugs by moving away from high dopamine D₂ receptor occupancy and limiting off-target neurotransmitter interactions causing adverse side-effects is likely to increase compliance (Li et al., 2016). There are a number of compounds in clinical trials at the moment for these unmet needs in schizophrenia treatment. Many of these drugs primarily act on serotonergic receptors and are aimed at either improving negative symptoms, treating cognitive impairments or addressing treatment-resistance associated with psychotic disorders (Fellner, 2017; Li et al., 2016). It remains to be seen whether these drugs will go on to be approved for clinical use.

Data collected here show that there has been a vast number of different pharmacological agents and other interventions tested in animal models of psychosis for the improvement of symptoms. One of the most obvious observations based on these data is that there is a substantial lack of replicability. Many of the compounds in the list collated were reported only a handful of times. In fact 62% of compounds were found to be reported only once in the literature. A lack of replicability has continued to be a major issue in all preclinical research fields. Good scientific practise that improves credibility and robustness of results is based heavily on replication of studies (Ioannidis, 2014). Faulty conclusions can impede further understanding of a concept and are wasteful in research. It is thought that one reason behind high rates of candidate drug failure in clinical trials is that they have not been based on robust experimental data to begin with (Steward, 2016). Not only is research which is not replicated not robust, but it is also arguably wasteful. With the high rate of false positive drugs identified to show efficacy in animal models, but not in later clinical trials (Moore, 2010), it seems that a lack of replication might play a role in translational failure of therapeutic targets from animal model studies to clinical trials in humans. It has been shown in recent years that multiple efforts to replicate findings in the literature have shown that many studies can't be replicated to the same extent as the original study (Ioannidis et al., 2014a), questioning the credibility of those initial findings. This is a major rate-limiting step in further understanding and the drug development process and I go on to explore further potential reasons behind this in the following chapters of this project.

During categorization, all effort was made to only record compounds thought to be given with the aim of reversing or attenuating the effects of a model induced.

However, the role of some compounds in experimental designs was not always clear. Results show that some pharmacological compounds can appear on both the list of interventions to induce a model and the list of treatment drugs tested in animal models of schizophrenia. For example, recent evidence shows that ketamine, a non-competitive NMDA glutamatergic receptor antagonist has been shown to have therapeutic efficacy as an antidepressant in the treatment of depression and bipolar disorder (Grady et al., 2017). It has also been found efficacious in patients who experience depressive symptoms in the context of psychotic symptoms (da Frola Ribeiro et al., 2016) or have a history of psychotic symptoms (Pennybaker et al., 2017). This is despite its wide use in the preclinical literature for the modelling of schizophrenia in animals as seen here, and its adverse effects when abused in humans (Li et al., 2011). It is thought that the time course of response after administration can produce differential effects, with psychotomimetic effects being seen first, followed by antidepressant effects of the drug (Duman et al., 2012). Similarly, nicotine is often used by schizophrenic individuals as a form of 'self-medication' due to its suggested ability of reducing psychiatric symptoms including cognitive deficits (Sacco et al., 2005) and side effects associated with antipsychotic treatment (Goff et al., 1992). However, nicotine is also a drug of abuse, which has been shown to increase dopamine release directly, similarly to other drugs of misuse (Brody et al., 2004). It has also been shown to be associated with an increased risk of psychosis and daily smokers have shown to have an earlier age of onset of psychotic disorder (Gurillo et al., 2015).

Ultimately, the development of novel antipsychotics in the field is made difficult not just by the limitations of animal models available, but also by the lack of a "gold standard" medication that is available in the clinic for the complete treatment of schizophrenia that could be used as a positive control for novel compounds (Jones et al., 2011). Of course, before we can consider the full therapeutic potential of some of these drugs, things such as dose required for efficacy should be considered as evidence shows that while these drugs may improve one domain of symptoms, it might also make another worse (Zajackowski et al., 2003). Moreover, external validity of experimental setups used is extremely important when it comes to finding treatment compounds that show efficacy in animals and also humans. Evidence shows that current antipsychotics have a small therapeutic window of efficacy before they produce unwanted behaviours that might affect overall performance of an individual in a behavioural test (Jones et al., 2011). External validity is a concept

further discussed in the context of developmental models of schizophrenia in Chapter 4.

3.4.4 Limitations

The search performed was a shallow, but broad overview meaning that search terms used were not specific and this means that studies could have been missed that would otherwise have been of relevance if they did not specifically mention “psychosis” or any other psychotic disorders as classified in DSM-5 or ICD-10 classification systems. This is further explored in Chapter 7.

The lists of models, treatments and outcomes were broadly validated based on information from the wider literature when compiling tables and checked overall by myself. Nevertheless, a small proportion of publications were outsourced for categorization to students working with our group at the time of performance and therefore there might be some discrepancy between what is considered a model, a treatment and a valid outcome to one investigator compared to another. Publications were also categorized by a single reviewer, potentially leading to some errors in categorization of publications due to human error or misinterpretation of a publication. For many publications it was often not clear what would be considered a model and what would be considered a treatment due to poor reporting of study intentions. I also found on many occasions that a study would be described to be of relevance to schizophrenia as well as multiple other psychiatric disorders, but there would be little further explanation to this association. In reflection of the heterogeneity seen in the reporting of model induction, outcome measurement and testing of compounds of potentially therapeutic value in models, I agree with previous calls in the literature for authors to “state the goals of their model” (Nestler and Hyman, 2010). It is recommended that authors more explicitly state the nature of their models and what specific symptoms they are modelling in order to provide more conceptual clarity and easier assessment of validity and utility of models by reviewers and readers in general (Nestler and Hyman, 2010). I would argue that in many cases the message of why a model was created, what it was expected to show and why certain outcome measures were used including what they were intended to measure and the purpose of testing the specific compounds reported was not always clear. I believe that this can make it often difficult to put a publication into context with other work going on, which could also be an issue behind limited translational potential of these studies.

4 Construct, Face, Predictive and External Validity: Developmental Animal Models of Schizophrenia

4.1 Introduction

The validity of animal models for a human condition is judged by the extent to which the features of the model and the condition it is expected to model are similar (Varga et al., 2010). The use of animal models in the research fields of uniquely human disorders, such as psychiatric disorders, has been a particularly difficult challenge (Powell and Miyakawa, 2006). In the previous chapter I discussed how vast the literature on animal models of schizophrenia is and reviewed different models reported to have relevance to the disorder. While there is no formal validation of models, the three validities most often discussed in the context of animal models of human conditions are construct, face and predictive validity. In the literature these validities are often variably defined, however here I will use definitions used in recent literature of animal models of psychiatric disorders (Belzung and Lemoine, 2011; Nestler and Hyman, 2010).

According to this, construct validity describes the degree of relevance a method used to construct a model has to the condition being modelled (Nestler and Hyman, 2010). In theory, to achieve construct validity we would create animal models of schizophrenia by mimicking the aetiology of the human disorder so that the animal would model, for example, neural or behavioural features of the disorder (Chadman et al., 2009). As the underlying pathophysiology and the exact aetiology has not been precisely established in the field of schizophrenia research, it is difficult to argue that animal models have high construct validity. This is especially because models based on “risk factors” are likely to also be of relevance to many other neurodevelopmental psychiatric disorders (Nestler and Hyman, 2010).

Face validity is the observed similarity between the pathophysiology of the animal modelling the human, and the human condition in question. This can include similarity in anatomical, biochemical, neuropathological and behavioural features between animal model and human disorder (Nestler and Hyman, 2010). This can be a very misleading measure of how valid an animal model is, especially as most models are different species to the one they are designed to model. It is therefore

rare for two species to exhibit the same behaviours even when the underlying biology is similar. In turn, similarities in behaviours do not translate to similar underlying causes in two different species (Geyer and Markou, 2000). Of course, it is also important to note that the validation of any animal model can only be as good as the information that is available from the clinical side of psychosis research (Geyer and Moghaddam, 2002). Face validity is therefore a very subjective measure of validity, and is a difficult concept to defend because many of the symptoms that define the human disorder being modelled are defined subjectively and diagnostic categories are continuously re-defined and changed (Geyer and Markou, 2000). For example, using guidelines such as the DSM, the same disorder can be characterized by opposite symptoms and many symptoms are not distinct from those of other psychiatric disorders (Donaldson and Hen, 2015; Weinberger, 2013). These limitations make it very difficult to create animals that are representative of the disorder in question.

Predictive validity is the extent to which a model is able to give accurate predictions about the efficacy of treatments in humans (Nestler and Hyman, 2010). In other words, to what degree does an animal model respond the same, as those with the human condition, to the same manipulation (Feifel and Shilling, 2010). As current antipsychotics are largely based on serendipitous clinical discoveries in the last century, the full predictive validity of animal models of psychotic disorders is not yet understood as they have not led to the development of any clinically approved treatment options for schizophrenia (Geyer et al., 2012). Geyer and Markou (2000) argue that because animal models exist solely for the purpose of bettering knowledge about a certain phenomenon, the only truly important criteria for evaluating an animal model of a human condition is its ability to have predictive validity (Geyer and Markou, 2000).

Of course, it is not only important for a model to be valid in terms of being able to recreate specific signs or psychological constructs of a disorder. In order to inform clinical trials for novel treatment development, treatment strategies have to be clinically relevant as well as similarly applied in humans as they have been in animals (van der Worp et al., 2010). Threats to external validity can include aspects such as a lack of generalisability – where, for example, animals tested do not represent the patient population being modelled. Differences in experimental design that make translation to the clinic unrealistic, such as unrealistic doses, timing of

administration, and timing of outcome assessment, can also affect the translation of results from animal models to clinical studies (van der Worp et al., 2010).

As mentioned, it is not thought that any one animal model is able to reiterate the complexity of the underlying biology of schizophrenia. Instead, many models model specific animal behaviours thought to be of relevance to human symptoms (Geyer and Moghaddam, 2002; Jones et al., 2011). It is very unlikely that animals will ever be able to model human-specific symptoms of schizophrenia such as altered perception, aberrant language or suicidal thoughts. However, not only do the measures taken from an animal model need to be reliable, but clinical measures must also be developed alongside these models that make it possible to produce meaningful inferences between the two (Geyer and Markou, 2000).

As seen in Chapter 3, one of the most prevalent and promising groups of animal models of schizophrenia involves the administration of environmental insults during development in animals to model etiological factors thought to play a role in the development of the disorder (Jones et al., 2011). As mentioned in the previous chapter these models are based on the neurodevelopmental hypothesis of schizophrenia, whereby early environmental insults during development are thought to interact with genetic predispositions to induce dysfunctions in neural systems that become apparent in later life (Fatemi and Folsom, 2009). As a result, these models are thought to be more promising than pharmacological models and in many cases genetic models too as they are able to provide more etiological validity for the human condition. They are able to recapitulate the delayed onset of symptoms seen in the clinic as well as create a biological abnormality that spans multiple neural systems thought to play a role in schizophrenia (Wilson et al., 2010).

Here I review this group of models and discuss their value to clinical research and novel drug development in the context of construct, face and predictive validity. I also look at the external validity of the experimental setups used.

4.2 Methods

Part of this work was carried out as part of a BSc Biological Sciences Honours dissertation project on isolation rearing and maternal separation models for which I would like to acknowledge Monica Dingwall (MD) for her work. I would also like to acknowledge our research assistants, Kaitlyn Hair (KH) and Paula Grill (PG), for their hard work on data extraction from other included studies. Inclusion and categorization of publications was carried out by myself. Final extracted data were checked and meta-analyses were run by myself.

4.2.1 Search strategy

Publications of relevance were filtered from the 3847 studies that had been screened for inclusion within this systematic review of animal models of psychotic disorders. They were filtered based on results from Phase II of the review, so that studies categorized as reporting either a developmental or a combination of a developmental and another type of model were considered to be of relevance. In light of the large corpus of data categorized as examples of developmental models, I only focused on three widely reported developmental models: animals infected prenatally in the womb, animals infected postnatally in early life and adversely reared animals.

4.2.2 Inclusion criteria

Inclusion and exclusion of publications occurred at full-text level in line with data extraction using inclusion/exclusion criteria specific for Phase III of the project in Chapter 2. To reiterate, publications were excluded from further analysis if they did not report 1) behavioural outcome measures, 2) an appropriate control, 3) data required for meta-analysis (i.e. number of animals used or SD/SEM).

4.2.3 Data extraction

Data extraction was carried out by a single reviewer (MD, PG, KH or myself). Only comparisons of developmental models or combination models where one of the models was considered a developmental model were extracted. Comparisons comparing control animals with model animals (termed model-characterising experiments) and those where model animals given a treatment to reverse the effect of the model were compared with animals of the same morbidity given no treatment

were included (treatment-exploring experiments), were both extracted, but analysed separately. Details of study characteristics extracted are specified in Chapter 2. Where the same outcome was reported in multiple ways within the same group of animals (e.g. locomotor activity reported in distance travelled and time spent moving), all outcome measures were extracted and nested before analyses.

4.2.4 Analysis

All methods used for analysis are described in detail in Chapter 2. Where the same outcome was reported in multiple ways, these measures were nested to give a single comparison within an experimental group for that outcome measure. Model-characterising comparisons and treatment-testing comparisons (see below for further detail) were meta-analysed separately. For specific models, treatments and outcome measures, and when appropriate and data sufficient (i.e. over 25 comparisons included in the meta-analysis), univariate meta-regression was performed to investigate potential sources of heterogeneity. This was done focusing on different components of study design characteristics, and a significance level of $p < 0.05$ was set for each test. To correct for multiplicity of testing, a Holm-Bonferroni adjusted critical p value was calculated to account for the number of variables tested within subgroup analyses. For most datasets, the adjusted critical p value was set at $p < 0.009$ for looking at the effect of 6 variables, with the exception of pre-pulse inhibition, where the effect of 7 variables was, explored making the adjusted critical p value to be set at $p < 0.007$. Heterogeneity is described using τ^2 (estimation of excess between-study variance), residual I^2 (the percentage of residual variation explained by between-study heterogeneity) and adjusted R^2 (adj R^2 ; the proportion of between-study variance explained by the model). Statistical analyses were performed using code written in R for combining data and Shiny Meta-analysis application (<https://qianying.shinyapps.io/Multi/>) for meta-analysis and univariate meta-regression.

For reference, a worse or improved behavioural outcome is used to describe how groups of animals perform on a behavioural measure compared to their control. For model characterising studies, we are comparing model animals to control, sham animals and therefore a worsening in outcome means the animal model of the disorder is performing poorer on the behaviour task compared to an unaffected animal.

For treatment testing studies, comparisons are made between two identical groups of animal models, where one is given a therapeutic intervention to improve their performance, while the other, control group, is not. Here, we would expect to see an improvement in outcome using a behavioural measure in animal models, which have been given the treatment, when compared to those animals, which have not been given the same treatment.

4.3 Results

4.3.1 Overview of the field and external validity of model studies

191 publications were identified from Phase II screening to include developmental manipulations, which induce behaviours thought to model human symptoms in schizophrenia as a result of prenatal or postnatal viral infection and adverse child rearing. During the data extraction process, 21 publications were excluded from this subset due to lack of data required for meta-analysis.

In total, 84 publications reporting 974 comparisons reported characterising animal models of schizophrenia by comparing developmentally disturbed animals to healthy control animals.

In total, model-characterising studies used a total of 21984 animals. Most experiments used rats (561 experiments, 12202 animals), with others using mice (371 experiments, 9241 animals) and monkeys (42 experiments, 541 animals). All of the monkeys used were rhesus macaques. The most commonly used strains of mouse were C57BL/6, of which most were specified to be a subline from the Jackson laboratory (124 comparisons, 3023 animals) and the rest were not further specified (61 comparisons, 1619 animals), and Balbc/c (40 comparisons, 1538 animals). 69 experiments did not state the strain of mouse used (1148 animals). The most commonly used rats were Wistar rats (292 comparisons, 4784 animals), Sprague Dawley rats (170 comparisons, 4330 animals), Fischer rats (49 comparisons, 910 animals) and Long-Evans rats (28 comparisons, 1359 animals).

Most experiments reported using male animals (579 comparisons, 11346 animals), with only 85 experiments using female animals (1909 animals). 288 experiments reported using both (7986 animals). 12 experiments were not clear on the sex of animals used in their experimental design, leaving 403 animals unaccounted for.

In addition, 17 publications reported 143 comparisons looked at testing the effect of various treatments in model animals by comparing treated to non-treated animals.

Treatment-testing studies used a total of 2483 animals. Most experiments used rats (84 experiments, 1222 animals), with others using mice (59 experiments, 1262 animals). The most commonly used strain of mouse was C57BL/6 (47 comparisons, 80% of all mice, 1047 animals). Rats used were most commonly Wistar rats (67 comparisons, 997 animals), and others were of the Sprague Dawley strain (17 comparisons, 225 animals). Most experiments reported using male animals (65 comparisons, 878 animals), with only 22 experiments exclusively using female animals (360 animals). Only 56 experiments reported using both sexes, however, these experiments reported using the largest number of animals (1246 animals).

The global estimate of effect of model-characterising studies was -0.68 SD units (95% CI -1.00- -0.36), meaning animal models of schizophrenia performed worse than control animals on behavioural measures by 0.68 SD units. The global estimate of efficacy of treatments administered in treatment-testing-studies was 1.01 SD units (95% CI 0.63-1.39), meaning treatments were able to improve behavioural outcomes in developmental models of schizophrenia by 1.01 SD units.

As development plays a key factor in schizophrenia and symptoms in humans arise at different stages, time of assessment was used to calculate stage of life of animal at measurement (Table 4.1).

Table 4.1 Developmental stages of life for mice, rats and rhesus macaques
Put together using the following sources: Sengupta (2013); Casey, Glatt & Lee (2015); <https://www.nc3rs.org.uk/macques/macques/life-history-and-diet/>.

	Mice	Rats	Rhesus Macaques
<i>Infant</i>	0-28 days	0-21 days	0-12 months
<i>Juvenile</i>	28-42 days	21-35 days	12-36 months
<i>Adolescent</i>	42-56 days	35-50 days	3-8 years
<i>Adult</i>	56+ days	50+ days	8+ years

In studies characterising the animal model induced, 13 groups of animals were measured as infants, 37 as juveniles, 20 as adolescents, and 186 as adults. When data were stratified according to these stages, we see that experiments measuring animals at the juvenile phase report a greater worsening in behaviour in comparison to animals being measured at all other stages of life (Table 4.2).

Table 4.2 Global estimates of effect of model on behaviour at specific stages of life
Estimates reported in SMD units, brackets contain 95% confidence intervals of this estimate and N correspond to the number of comparisons contributing to the

	Infant	Juvenile	Adolescent	Adult	Global
Effect size (95% CI)	-0.60 SD (-1.00;-0.20)	-3.02 SD (-7.65;1.62)	-0.53 SD (-0.70;-0.37)	-0.746 SD (-1.12;-0.37)	-0.68 SD (-1.00;-0.36)
N	13	37	20	186	256

Combining infant, juvenile and adolescent animals into one group to include all measurements taken in young animals before puberty still shows that overall behaviour is worse in younger animals than in animals measured at a later stage in life (Table 4.3). Even so, it is obvious that most studies measure behaviour in adult animals.

Table 4.3 Global estimates of effect of model on behaviour in young and adult animals

Estimates reported in SMD units, brackets contain 95% confidence intervals of this estimate and N correspond to the number of comparisons contributing to the

	Young	Adult	Global
Effect size (95% CI)	-1.47 SD (-3.50;0.56)	-0.746 SD (-1.12;-0.37)	-0.68 SD (-1.00;-0.36)
N	70	186	256

Treatments are also mainly tested in adults, with only two groups of animals being measured at the juvenile phase of their life. For this reason, it is difficult to analyse how well treatments work at different stages of an animal model's life.

Table 4.4 Global estimates of efficacy for treatments in young and adult animals

Estimates reported in SMD units, brackets contain 95% confidence intervals of this estimate and N correspond to the number of comparisons contributing to the calculation, nested where appropriate.

	Young	Adult	Global
Effect size (95% CI)	Data not sufficient	1.03 SD (0.64;1.42)	1.01 SD (0.63;1.39)
N		57	59

4.3.2 Exploring construct validity

Construct validity describes the degree of similarity between an animal model and the human condition it is intended to model in terms of underlying neurobiological mechanisms (van der Staay et al., 2009). In order to explore the construct validity of developmentally induced animal models of schizophrenia further, I looked at the way in which models were induced and what effect these different experimental setups had on behaviour in these animals. Overall, model-characterising studies reported using 37 different methods to induce animal models of schizophrenia (Figure 4.1). 12 of these induction methods were combination models where the main model was an insult to the animal's development through prenatal or postnatal disturbance to normal development, combined with a second hit usually in later life involving some form of stress to the animal or pharmacological hit to exacerbate the underlying condition. The most common methods used to induce the model were by infection of mothers during pregnancy with the virus polyinosinic:polycytidylic acid (poly I:C) (433 comparisons, 11658 animals), the neurotoxin methylazoxymethanol acetate (MAM) (164 comparisons, 2646 animals) and lipopolysaccharide (LPS), a toxic component in gram-negative bacteria (98 comparisons, 2246 animals).

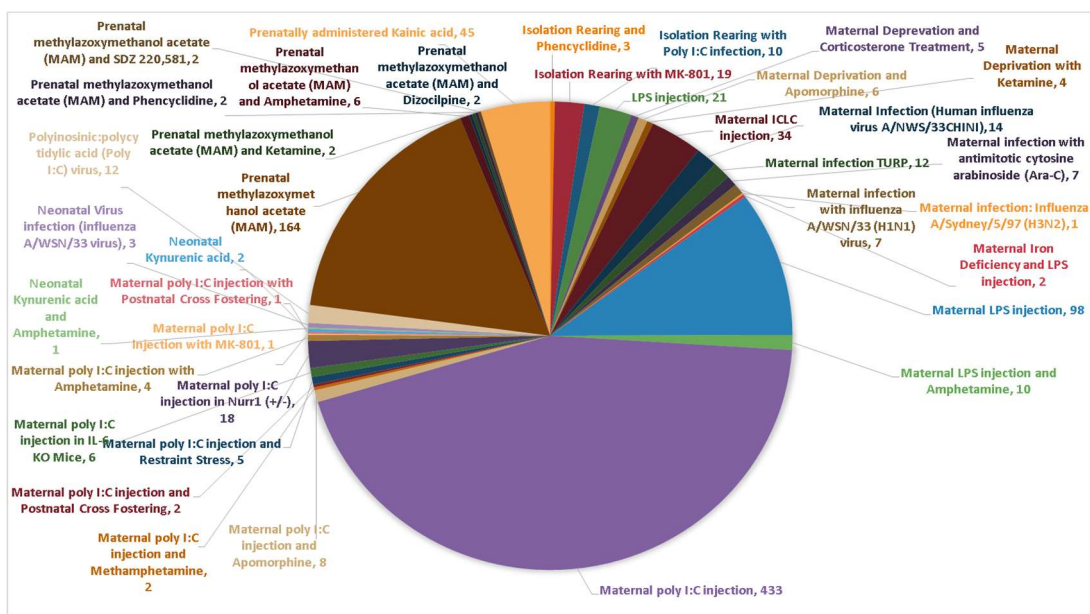


Figure 4.1 Developmental methods used to induce psychosis under models of infection and adverse rearing conditions

In further analysis, for the sake of simplicity and clarity these models were grouped according to the type of method used to induce the model: prenatal insult, postnatal insult, adverse rearing condition, and combination model. These methods are now reviewed here in more detail.

4.3.2.1 Prenatal infection

896 experiments induced their model by affecting development before birth. This included the following classic maternal immune activation models: injections of LPS, TURP, poly I:C, MAM, ICLC (i.e. modified form of poly I:C), kainic acid, cytosine arabinoside (Ara-C), or various strains of influenza virus. Some of these interventions were combined with each other as well as with other maternal insults such as maternal iron deficiency. Many of these prenatal models were also combined with a second hit of pharmacological interventions administered later in the life of the pups including amphetamine, methamphetamine, apomorphine, and NMDA antagonists such as SDZ 220,581, ketamine, phencyclidine, or dizocilpine. Developmental models were also combined with genetic models, for example, poly I:C injection was given to IL-6 KO mice or Nurr1 (+/-) animals. Some animal models also combined a developmental intervention with an environmental stressor later in life such as juvenile restraint stress or postnatal cross fostering.

Most of the animals reported to be subjected to this method of model induction were mice (356 experiments, 8992 animals) or rats (498 experiments, 11282 animals), with 42 comparisons carried out in monkeys (42 experiments, 541 animals). All monkeys were rhesus monkeys. The most commonly used strains were C57BL6/J for mice (124 experiments, 3023 animals) and Sprague Dawley for rats (147 experiments, 3966 animals). Studies using these methods of model induction reported an overwhelming number of comparisons carried out in male animals (509 experiments, 10275 animals), but many comparisons were also found to be among groups of animals of both sexes (287 experiments, 7969 animals). 78 comparisons were reported in only female animals (8.7% of all prenatal infection experiments, 1827 animals). After birth, animals were measured at a number of different stages of life, including 30 measurements taken from infant animals (623 animals), 219 measurements taken from juvenile animals (3301 animals), 76 measurements taken from adolescent animals (3076 animals) and 571 measurements taken from adult animals (13815 animals). Other details about experimental design are shown in Table 4.5 for each type of model used.

Table 4.5 Details of experiments describing prenatal infection models

Model Induction Method	Dose to induce model	# of Admin	Age at first Admin	Second-hit	Dose	Total # of Admins	Admins per day	Age at Admin	Number of comparisons	Number of animals used		
Maternal ICLC injection	0.25 mg/kg	3	GD 43						17	196		
			GD 100						17	213		
Maternal Infection (Human influenza virus A/NWS/33CHINI)	10 mg/kg 20 mg/kg 10000 pfu 6000 pfu	1	GD 9.5						2	30		
			GD 9.5						4	60		
			GD 9.5						7	245		
			GD 9.5						8	324		
Maternal infection TURP	0.05 mg/kg 0.75 mg/kg 10 µL	2	GD 10						4	59		
			GD 18						3	45		
			GD 10						3	54		
			GD 18						2	36		
Maternal infection with Ara-C	30 mg/kg	2	GD 19					7	280			
Maternal infection: Influenza A/Sydney/5/97 (H3N2)	1 ml	1	GD 119					1	19			
Maternal Iron Deficiency and LPS injection	-	27	GD 2	Juvenile LPS injection	50 ug/kg	2	1	-7	2	23		
Maternal LPS injection	0.004 mg/kg 0.025 mg/kg 0.05 mg/kg 0.1 mg/kg 0.25 mg/kg 0.75 mg/kg 1 mg/kg 2 mg/kg 10 µL 100 mg/kg 1 mg/kg 0.05 mg/kg 0.5 mg/kg 1 mg/kg 2 mg/kg 4 mg/kg 5 mg/kg 10 mg/kg 10 µL 20 mg/kg 60 mg/kg 6000 pfu 5 mg/kg 5 mg/kg 2 mg/kg 5 mg/kg 2 mg/kg 20 mg/kg 20 mg/kg 5 mg/kg 5 mg/kg 5 mg/kg	2	GD 125							7	114	
			GD 10							4	59	
			GD 18								1	20
			GD 15								4	59
			GD 15								10	340
			GD 18								1	15
			GD 15								2	27
			GD 7								28	472
			GD 1								18	324
			GD 1								1	50
			GD 2								1	14
			GD 3								1	14
			GD 4								1	50
			GD 5								1	30
			GD 6								1	14
			GD 7								1	50
			GD 8								1	14
			GD 9								1	18
			GD 10								1	18
			GD 11								1	30
			GD 12								1	30
			GD 13								1	14
			GD 14								1	18
			GD 15								1	14
			GD 22								20	618
			GD 10								1	18
			GD 15								1	18
GD 18								2	36			
GD 15								3	112			
GD 2								2	17			
GD 7	Amphetamine	8	GD 7		0.5 mg/kg				10	192		
Maternal poly I:C injection	0.05 mg/kg 0.5 mg/kg 1 mg/kg 2 mg/kg 4 mg/kg 5 mg/kg 10 mg/kg 10 µL 20 mg/kg 60 mg/kg 6000 pfu 5 mg/kg 5 mg/kg 2 mg/kg 5 mg/kg 2 mg/kg 20 mg/kg 20 mg/kg 5 mg/kg 5 mg/kg 5 mg/kg	2	GD 18							3	45	
			GD 20							14	322	
			GD 10								4	59
			GD 15								2	32
			GD 9								15	764
			GD 14								42	773
			GD 15								108	3745
			GD 16								1	12
			GD 17								4	32
			GD 18								4	72
			GD 6								3	88
			GD 9								99	1870
			GD 13								3	95
			GD 16								4	128
			GD 17								61	1421
			GD 13								2	140
			GD 18								1	62
			GD 12								6	880
			GD 9.5								2	30
			GD 15								3	54
			GD 12								5	75
			GD 13								11	461
			GD 14								3	69
			GD 9.5								3	69
			GD 11								1	29
			GD 9.5								4	134
			GD 17	Apomorphine	1	GD 17		5 mg/kg	1			6
GD 9	Apomorphine	1	GD 9		0.75 mg/kg	1	1	35	1	16		
								70	1	20		
GD 12	Methamphetamine	6	GD 12		2 mg/kg	1	1		2	80		
GD 9	Amphetamine	1	GD 9		2.5 mg/kg				1	19		
GD 9	Amphetamine	1	GD 9		2.5 mg/kg	1	1	92	3	42		
GD 9	MK-801	1	GD 9		0.1 mg/kg				1	19		
GD 12	Restraint Stress	1	GD 12		2 hrs/day	3	1	34	5	75		
GD 13	IL-6 KO Mice	1	GD 13						6	272		
GD 6	Nurr1 genetic knockout	1	GD 6						9	120		
GD 9	Nurr1 genetic knockout	1	GD 9						9	120		
GD 9	Postnatal Cross Fostering	1	GD 9						3	27		

Model Induction Method	Dose to induce model	# of Admin	Age at first Admin	Second-hit	Dose	Total # of Admins	Admins per day	Age at Admin	Number of comparisons	Number of animals used	
Prenatal methylazoxymethanol acetate (MAM)	20 mg/ml	1	GD 13						2	20	
			GD 12						2	20	
			GD 14						2	20	
			GD 15						2	20	
	20 mg/kg	1	GD 9							27	304
			GD 10							27	322
			GD 11							25	299
			GD 12							27	298
			GD 17							1	10
	22 mg/kg	1	GD 17						16	242	
	22 mg/ml	1	GD 17						12	288	
	25 mg/kg	1	GD 17						21	803	
	20 mg/kg	1	GD 17	Amphetamine	0.5 mg/kg	1	1	2350	1	23	
	22 mg/kg	1	GD 17	Amphetamine	2 mg/kg	1	1	1416	5	280	
Dizocilpine (MK801)				0.1 mg/kg	1	1	96	2	24		
Ketamine				10 mg/kg	1	1	96	2	24		
Phencyclidine				2.5 mg/kg	1	1	96	2	24		
SDZ 220.581				3 mg/kg	1	1	96	2	24		
Prenatally administered Kainic acid	0.2 µL	1	GD 18					45	910		

Overall, behaviour in animals where the model was induced through adverse rearing conditions worsened outcome by 0.74 SD units (95% CI -1.15--0.34) in model animals, compared to healthy, control animals ($p < 0.0001$, $n = 242$ comparisons). Substantial heterogeneity was observed in the data ($\tau^2 = 2.4092$, $I^2 = 97.5\%$), which was in part explained by the number of times intervention was administered prenatally in the mother ($p = 0.0006$, $\tau^2 = 2.156$, $I^2 = 99.75\%$, $\text{adj } R^2 = 10.53\%$, Figure 4.2). Heterogeneity was not significantly explained by any of the other study characteristic variables investigated with univariate meta-regression, namely: animal, strain and sex of animals, method used to induce animal model, stage of life behaviour was measured at, time of outcome assessment, dose administered in mother to induce the model and time of administration during gestation.

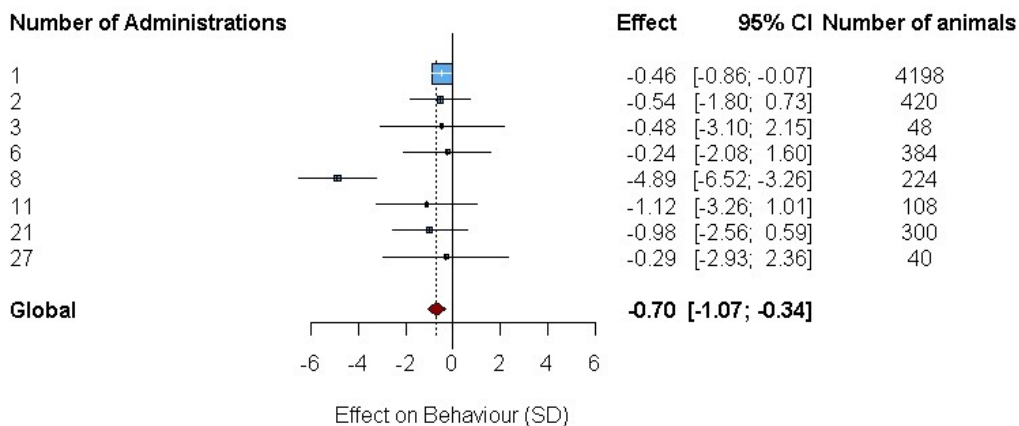


Figure 4.2 Relationship between the number of administrations of prenatal infection and reported effect size in model-characterising studies

4.3.2.2 Postnatal infection

31 experiments using 501 animals induced their model by affecting development soon after birth. This included the following induction methods: neonatal kynurenic acid, neonatal kynurenic acid and amphetamine, neonatal viral infection using the influenza A/WSN/33 virus, and the poly I:C virus. Early in development meant administering these interventions at any time from postnatal day 2 up until postnatal day 14.

Animals subjected to this method of model induction were rats (16 experiments, using 252 animals) or mice (15 experiments, using 249 animals). The rats were of the strains Sprague Dawley (4 experiments, using 60 animals) or Wistar (12 experiments, using 192 animals). The mice used were C57BL/6 mice (6 experiments, using 109 animals) or ICR mice (9 experiments, using 140 animals). Experiments overwhelmingly reported exclusively using male animals in their design (26 experiments, 84% of all experiments using postnatal models, 424 animals). One experiment reported using both male and female animals (17 animals), and 4 reported using only female animals within their study design (60 animals). Once models were established, outcome was usually measured in adulthood (27 experiments, 437 animals), with only a few studies measuring outcome in adolescence (4 experiments, 64 animals). Models using simple neonatal kynurenic acid treatment administered 20 repeated dosages of 200 mg/kg during development (2 experiments, 48 animals). When this was combined with amphetamine, an injection was only given once during development (1 experiment, 8 animals). Neonatal infection using a strain of the influenza virus was either administered once at 2400 plaque forming units (1 experiment, 17 animals) or repeatedly during development at a dosage of 200 mg/kg (2 experiments, 36 animals). Experiments reporting early postnatal infection using the Poly I:C virus to create their model (25 experiments) used a variety of experimental paradigms. In four experiments the virus was administered once during development (60 animals), in 12 experiments it was administered three times during development (192 animals) and in 9 experiments it was administered repeatedly 5 times during development (140 animals). The dosage of the virus varied from 2-5 mg/kg.

Overall, behaviour in animals where the model was induced through postnatal infection worsened outcome by 1.07 SD units (95% CI -1.52--0.61) in these animals, compared to healthy, control animals ($p < 0.0001$, $n = 12$ comparisons). Substantial

heterogeneity was observed in the data ($\tau^2 = 0.45$, $I^2 = 94.2\%$), however, due to low sample size after nesting of data, I was unfortunately unable to further explore sources of this heterogeneity.

4.3.2.3 Adverse rearing

48 experiments, using a total of 678 animals induced their model by affecting development during rearing of animals. This included the following induction methods: isolation rearing and phencyclidine, isolation rearing and MK-801, maternal deprivation and corticosterone treatment, maternal deprivation and apomorphine, maternal deprivation and ketamine, isolation rearing and poly I:C infection, and maternal poly I:C injection and postnatal cross fostering.

Overall most of the animals subjected to this method of model induction were rats, with all but one experiment using rats in their experimental design (47 experiments, 669 animals). One experiment reported using mice of the strain C57BL6/J (9 animals). The rats used were of a variety of different strains, including Fischer rats (2 experiments, 32 animals), Hooded Lister rats (3 experiments, 23 animals), Lewis rats (2 experiments, 28 animals), Sprague Dawley rats (19 experiments, 305 animals) and Wistar rats (21 experiments, 282 animals), with the latter two strains being used most widely across experiments. The sex of the animals was mostly male (44 experiments, 646 animals), with 3 experiments reporting the use of female animals (23 animals) and one experiment reportedly using both sexes (9 animals). Those experiments utilising isolation rearing as a model reported isolating animals for 5 to just over 12 weeks, always starting at 3 weeks of age. Animals were most commonly isolated for either 8 (12 experiments, 38% of all isolation rearing experiments, 183 animals) or 12 weeks (10 experiments, 31 % of all isolation rearing experiments, 120 animals). Experimental designs where the pup was deprived from the mother and separated from her were reported to always last for 24 hours. Behaviour in animals was usually measured at the adult phase of life (43 experiments, 638 animals), but 5 experiments also measured behaviours at the juvenile stage of life (40 animals). Where pharmacological agents were administered as a second hit to the model, these were usually administered just before behaviour was assessed.

Overall, behaviour in animals where the model was induced through adverse rearing conditions worsened outcome by 0.37 SD units (95% CI -0.63--0.11, compared to other models: Table 4.6) in these animals, compared to healthy, control animals ($p < 0.0001$, $n = 20$ comparisons). Substantial heterogeneity was observed in the data ($\tau^2 = 0.2766$, $I^2 = 91.9\%$), however, due to low sample size after nesting of data, I was unfortunately unable to further explore sources of this heterogeneity.

Table 4.6 Global estimates of efficacy for methods used to induce the animal model
Estimates are reported in SMD units, brackets contain 95% confidence intervals of this estimate and N correspond to the number of comparisons contributing to the calculation, nested where appropriate.

	Adverse Rearing	Prenatal Infection	Postnatal Infection	Global
Effect size (95% CI)	-0.37 SD (-0.63;-0.11)	-0.70 SD (-1.07;-0.34)	-1.07 SD (-1.52;-0.61)	-0.68 SD (-1.00;-0.36)
N	20	242	12	256

4.3.2.4 Combined or Simple Models

853 experimental comparisons involved using simple animal models of schizophrenia to characterise the model. Combined models were less prevalent with only 121 experiments using an experimental design where a second-hit was used to create the model. When stratifying the dataset by whether experiments used simple or combined models as part of their experimental design, there was little difference between the two types of studies and their effect on behaviour in animal models of schizophrenia when compared to control animals. Overall, simple models worsened behavioural outcome by 0.70 SD units (95% CI -1.08—0.31) in model animals when compared to control animals ($p < 0.0001$, $n = 225$ comparisons). On the other hand, combined models worsened behavioural outcome by 0.68 SD units (95% CI -1.00—0.36) in model animals when compared to control animals ($p < 0.0001$, $n = 48$ comparisons).

Table 4.7 Global estimates of effect for simple or combined model induction methods
Estimates are reported in SMD units, brackets contain 95% confidence intervals of this estimate and N correspond to the number of comparisons contributing to the calculation,

	Simple Models	Combined Models	Global
Effect size (95% CI)	-0.70 SD (-1.08;-0.31)	-0.68 SD (-1.00;-0.36)	-0.68 SD (-1.00;-0.36)
N	225	48	256

4.3.3 Exploring predictive validity

Predictive validity of an animal model is the extent to which phenotypic effects in the model are similar to those in humans in response to the same therapeutic intervention (Kumar et al., 2016). Therefore, predictive validity expects the model to be able to predict behaviour in the disorder that it has been designed to model. This is especially important in novel treatment development (van der Staay et al., 2009). For example, if an animal model is able to correctly predict the efficacy of a treatment drug in the human condition, then we can say that it has high predictive validity. In order to explore the potential predictive validity of developmental animal models of schizophrenia, I looked at the effect of different treatments administered in studies reporting these animals.

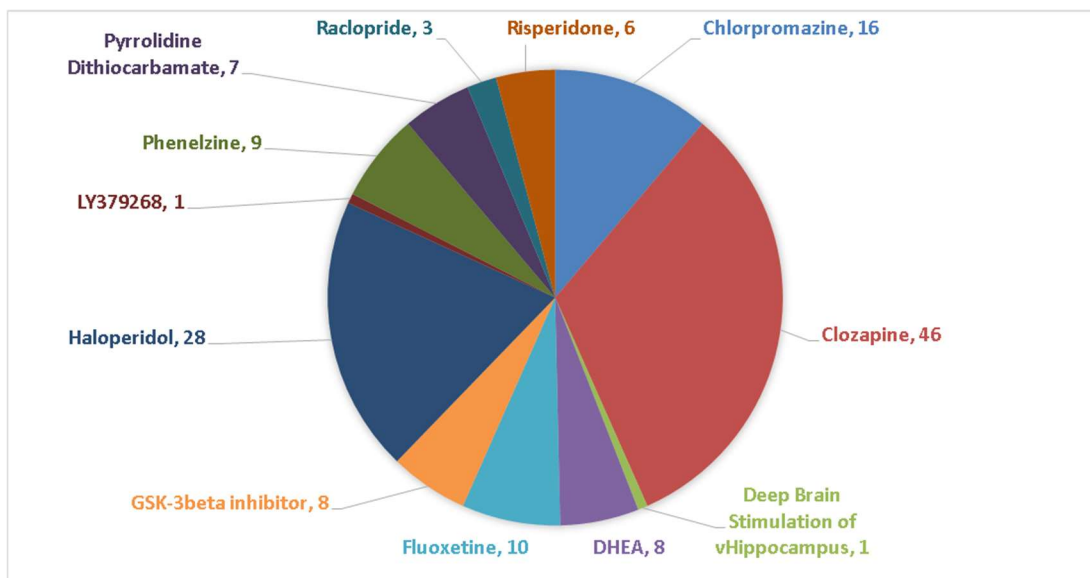


Figure 4.3 Treatments recorded to be tested in the literature. Number show number of experiments exploring these treatments.

142 comparisons of treatment-testing experiments were identified that compare a group of model animals with another group of model animals who have been given a potential therapeutic intervention to reverse or alleviate outcome in behaviour. 12 different interventions were tested in this subset of the literature (Figure 4.3). The most common of these were the already clinically established anti-psychotics clozapine, haloperidol and chlorpromazine.

It would have been interesting to compare experiments using established treatments with those using potential treatments, however, these newer compounds were reported in a few experiments and it was therefore not possible to make accurate comparisons. Here, I review the three most common treatments administered in the literature reviewed in further detail to explore the possible predictive validity of the models discussed in this chapter.

4.3.3.1 Clozapine

46 experiments using 815 animals tested the effects of clozapine in an animal model of developmentally-induced schizophrenia. Clozapine was administered to animal models of prenatal LPS or poly I:C infection, or was given to animals infected with poly I:C at an early stage of postnatal development.

The animals used to test the effects of the drug were a mixture of rats (30 experiments, 454 animals) and mice (16 experiments, 361 animals). The rats used were all of the Wistar strain. The mice used were Balb/c mice (2 experiments, 36 animals) or C57BL6 mice (14 experiments, 325 animals). Experiments mostly used male animals (24 experiments, 352 animals), with 13 comparisons including animals of both sexes (307 animals) and 9 using only female animals (156 animals). Dosage of the drug administered varied from 1 mg/kg to 25 mg/kg. Most experiments reported administering clozapine at a dose somewhere in the middle of this range at doses of 5 mg/kg (12 experiments, 230 animals), 10 mg/kg (16 experiments, 216 animals) or 15 mg/kg (9 experiments, 232 animals). The effect of all treatments were measured in adulthood after either a single treatment of clozapine or after 3-30 repeated administrations of the drug. Most experiments either administered the drug once (17 experiments) or 8 times (16 experiments) before behavioural assessment.

Overall, clozapine improved behaviour in animal models of schizophrenia by 1.29 SD units (95% CI -0.61-1.98), when compared to untreated model animals ($p < 0.001$, $n = 23$ comparisons). Substantial heterogeneity was observed in the data ($\tau^2 = 2.39$, $I^2 = 98.6\%$), but due to the low sample size once comparisons in the same group were nested, I was not able to explore this any further.

4.3.3.2 Haloperidol

28 experiments using 656 animals tested the effects of haloperidol in an animal model of developmentally-induced schizophrenia. Treatment was tested in animal models induced using prenatal infection with LPS or the poly I:C virus.

Haloperidol was reported in studies to be tested mostly in mice (21 experiments, 505 animals), with a few studies testing the effects of the drug in rats (7 experiments, 152 animals). All rats used were of the Wistar strain and the mice used were mainly of the C57BL/6 strain (19 experiments, 469 animals), with two comparisons being reported in Balb/c mice (36 animals). Interestingly most treatments were tested in groups of animals that included both male and female animals (21 experiments, 512 animals), with only three comparisons being reported in only male animals (72 animals), and 3 reported in only female animals (72 animals). The effect of treatment was always measured in adulthood in response to either 0.1 mg/kg (7 experiments), 3 mg/kg (19 experiments), or 5 mg/kg (2 experiments) of the drug. Usually haloperidol was given to animals only once before testing its effect (20 experiments), however, two experiments reported measuring behaviour after six repeated administrations of the drug and 6 experiments reported measuring behaviour after 11 repeated administrations of the drug.

Overall, administration of haloperidol led to an improvement in behaviour in animal models of schizophrenia by 0.80 SD units (95% CI -0.65-2.25) when compared to untreated model animals ($p < 0.001$, $n = 10$ comparisons). Substantial heterogeneity was observed in the data ($\tau^2 = 4.04$, $I^2 = 99.3\%$), but due to low sample size once comparisons in the same group had been nested, I was not able to explore this any further.

4.3.3.3 Chlorpromazine

16 experiments using 200 animals tested the effects of chlorpromazine in an animal model of prenatal infection with LPS. As all experiments came from the same lab, there was little variability in study design. All of the animals used to test the effects of the drug were Wistar rats (16 experiments, 252 animals). 10 of the comparisons reported in the literature used male animals (124 animals) and six used female animals (76 animals). Treatments were always tested in adulthood and behaviour was analysed in response to 8 repeated treatments of 10mg/kg of the drug.

Overall, behaviour in animal models of schizophrenia was improved by chlorpromazine by 1.17 SD units (95% CI -0.24-2.58, compared to other treatment drugs: Table 4.8), when compared to untreated model animals ($p < 0.001$, $n = 7$ comparisons). Substantial heterogeneity was observed in the data ($\tau^2 = 2.24$, $I^2 = 96.8\%$), but due to low sample size I was not able to explore this any further.

Table 4.8 Global estimates of treatment efficacy in developmentally-induced animal models of schizophrenia
 Estimates are reported in SMD units, brackets contain 95% confidence intervals of this estimate and N correspond to the number of comparisons contributing to the calculation, nested where appropriate.

	Clozapine	Haloperidol	Chlorpromazine	Global
Effect size (95% CI)	1.29 SD (0.61;1.98)	0.80 SD (-0.65;2.25)	1.17 SD (-0.24;2.58)	1.01 SD (0.63;1.39)
N	23	10	7	59

4.3.4 Exploring face validity

Face validity of an animal model looks at the degree to which we can see a descriptive similarity between the model and those that are being modelled. For example, do we see a behavioural dysfunction in the animal model in question similar to that in the human who is affected by the disorder we are attempting to model (van der Staay et al., 2009). To explore the face validity of developmental models of schizophrenia, I looked at different outcome measures that were used to measure behaviour in these animals.

4.3.4.1 Model-characterising studies

Altogether 32 different outcome measures were tested in the developmental models of schizophrenia reviewed here (Figure 4.4). The same outcomes performed in the same group of animals were nested for meta-analysis. The most commonly reported outcomes were pre-pulse inhibition (357 comparisons, 9141 animals), social interaction (153 comparisons, 2405 animals) and locomotor activity (107 comparisons, 2269 animals). These are further explored below.

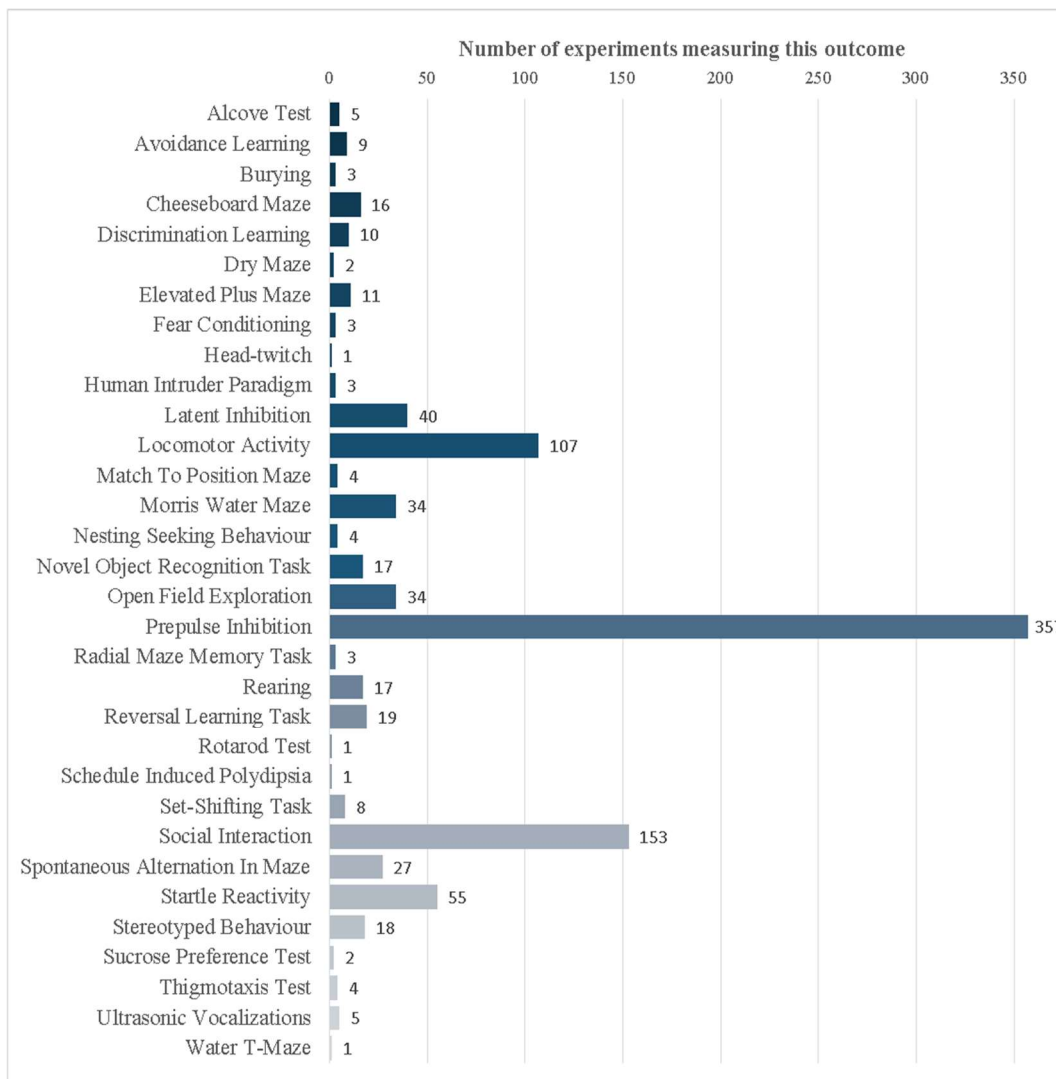


Figure 4.4 Prevalence of behavioural outcome measures reported model-characterising experiments of developmentally-induced animal models of schizophrenia

4.3.4.1.1 Pre-pulse inhibition

Pre-pulse inhibition (PPI) is a neurological phenomenon where the protective response to a startling stimulus is weakened when it is preceded by a weaker stimulus that is non-startling (Valsamis and Schmid, 2011). Here the pulse is the starting stimulus and the weaker pre-stimulus is referred to as the pre-pulse.

357 comparisons of PPI were identified within the model-characterising studies included in this analysis. Most of these measurements were performed on rats (246 comparisons, 5423 animals) and mice (174 comparisons, 4267 animals), with only six comparisons being measured in monkeys (72 animals). The rats used were most

commonly Sprague Dawley rats (102 comparisons, 2076 animals) and Wistar rats (122 comparisons, 2588 animals). The strain of mouse used was much less consistent across different studies, but a large proportion of studies used either C57BL/6 mice (40 comparisons, 1216 animals) or C57BL/6J mice (47 comparisons, 1025 animals). Most studies used either only male animals (241 comparisons, 4601 animals) or both male and female animals within the same experiment (110 comparisons, 3366 animals). Only 55 experiments reported using only female animals (1117 animals) and for 10 experiments it was unclear which sex of animals was used (339 animals). Most experiments measured pre-pulse inhibition when the animal was an adult (323 comparisons, using 7089 animals). Seven experiments measured pre-pulse inhibition when the animal was an infant (168 animals), 52 experiments when the animal was a juvenile (961), and 44 experiments did so during adolescence (1544 animals). Background noise used during experiments ranged from 45-72 decibels (dB), with the most common levels of white noise used being 65 Db (179 comparisons, 42% of all PPI experiments, 3900 animals) and 46 Db (32 comparisons, 8% of all PPI experiments, 852 animals). Startling pulses reported by studies ranged from 65 Db to 120 Db, with most experiments using either 120 Db (187 experiments, 4157 animals) or 100 Db (59 experiments, 1277 animals). The pre-pulse used ranged from 1-67 Db above background noise. Most commonly, pre-pulse was reported as being 12 Db (46 comparisons, 914 animals), 10 Db (30 comparisons, 778 animals) or six Db (41 comparisons, 819 animals) above the background noise. Ten experiments used a visual stimulus as a pre-pulse. Overall, most studies measured PPI at a range of different conditions. Unfortunately, 15 experiments did not specify the background noise level used, 99 experiments did not specify the startling pulse strength used, and 154 experiments were unclear about what type of pre-pulse was used before the startling stimulus.

Overall, behaviour measured using PPI was worse by 0.65 SD units (95% CI -0.54--0.76) in model animals compared to healthy control animals ($p < 0.0001$, $n = 268$ comparisons). Substantial heterogeneity was observed in the data ($\tau^2 = 0.7589$, $I^2 = 97.9\%$), which was in part explained by the stage of life at which animals were measured ($p = 0.0015$, $\tau^2 = 0.7178$, $I^2 = 99.07\%$, $\text{adj } R^2 = 5.41\%$, Figure 4.5) and, stemming from this, the time of outcome assessment ($p = 0.0003$, $\tau^2 = 0.7170$, $I^2 = 99.08\%$, $\text{adj } R^2 = 5.52\%$, Figure 4.6). Heterogeneity was not explained by any of the other study characteristic variables investigated with univariate meta-regression,

namely: animal, strain and sex of animals, method used to induce animal model and pulse strength above background noise.

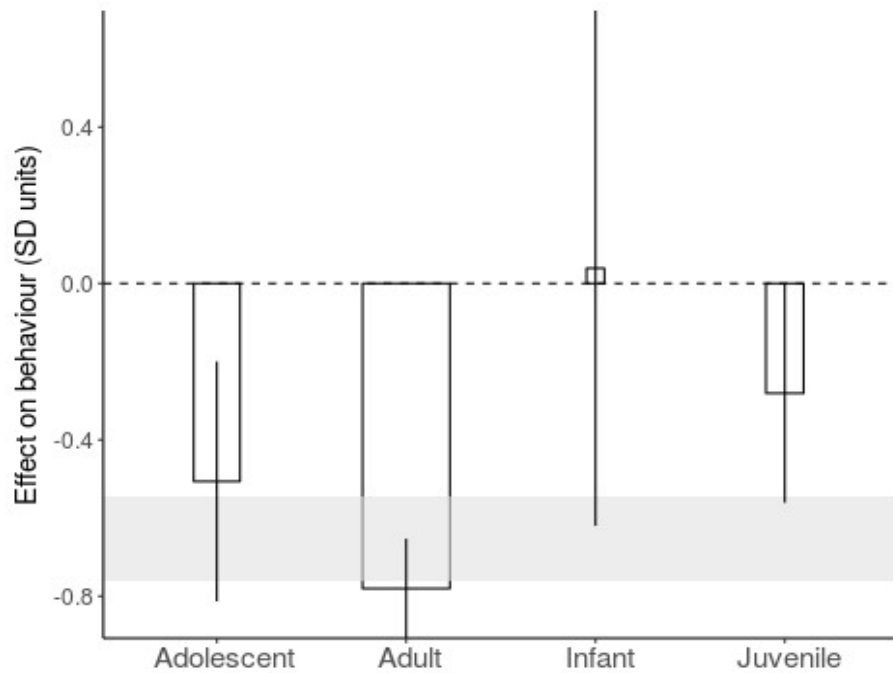


Figure 4.5 Relationship between stage of life at which behaviour is measured in animals and reported effect size in model-characterising

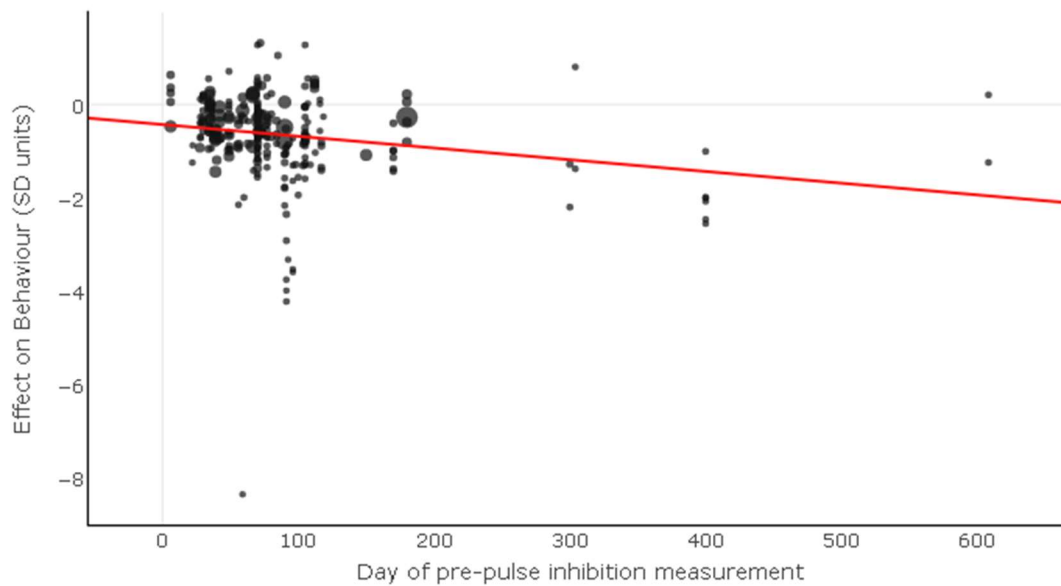


Figure 4.6 Relationship between time of outcome assessment and reported effect size in model-characterising experiments

4.3.4.1.2 Social interaction

Social interaction is measured in animals by quantifying behaviours that are observed during the interaction of two or more individuals of the same species, and can be measured by a variety of different measures (Wilson and Koenig, 2014). These behavioural measures include playful and aggressive acts of interaction in a normal environment. They are thought to show relevance to sociability in humans, and thus can be of relevance to negative symptoms in schizophrenia such as asociality.

I identified 153 individual comparisons measuring social interaction between model and control animals in the included dataset, using 2405 animals. Most of the animals used were rats (120 experiments, 1913 animals) or monkeys (28 experiments, 336 animals). Only 5 experiments used mice (156 animals). All of the monkeys used were Rhesus macaques, while rats were either Wistar (91 experiments, 1235 animals), Sprague Dawley (9 experiments, 338 animals), or Fischer (20 experiments, 340 animals) rats. The mice used were either of the strains Balb/c (2 experiments, 76 animals), C57BL/6 (2 experiments, 64 animals) or ICR (1 experiment, 16 animals). Animals used within experiments were either exclusively male (116 experiments, 1759 animals) or a combination of male and female (32 experiments, 476 animals). Five experiments did not state the sex of the animals used. Social interaction, unlike other outcome measures, was usually assessed in juveniles (113 experiments, 1323 animals). 30 experiments measured social interaction in adult animals (746 animals), 8 measured this behaviour in adolescence (312 animals) and two measured it in infancy (24 animals).

Overall, measuring social interaction in developmental animal models of schizophrenia showed a worsening in behaviour by 0.61 SD units (95% CI -0.86-0.36) when compared to healthy, control animals ($p < 0.0001$, $n = 19$ comparisons). Substantial heterogeneity was observed in the data ($\tau^2 = 0.2508$, $I^2 = 97.6\%$), however, due to low sample size after nesting of data, I was unable to further explore sources of this heterogeneity.

4.3.4.1.3 Locomotor activity

I identified 107 individual comparisons measuring locomotor activity in model-characterising studies, using 2269 animals. These experiments mainly reported

using rats (70 experiments, 1484 animals) and mice (26 experiments, 767 animals) to measure locomotor activity, with only one experiment measuring the behaviour in monkeys (19 animals). As with studies measuring PPI, the strains most commonly used were Sprague Dawley (26 experiments, 551 animals) and Wistar (24 experiments, 349 animals) for rats and C57BL/6J for mice (19 experiments, 343 animals). Most experiments either used male animals (63 experiments, 1267 animals) or reported using a combination of both male and female animals (33 experiments, 768 animals). Ten experiments reported using female animals (10 experiments, 187 animals) and one experiment did not specify the sex of animals used in its experimental design, leaving 48 animals unaccounted for. Locomotor activity was usually assessed in adulthood of the animals (74 comparisons, 1546 animals), but was also frequently measured at the juvenile phase (22 experiments, 301 animals). Much fewer experiments reported measuring locomotor activity during infancy (3 experiments, 93 animals) or during adolescence (8 experiments, 330 animals).

Overall, measuring locomotor activity in developmental animal models of schizophrenia showed a worsening in behaviour by 7.97 SD units (95% CI -16.41-0.47, compared to other outcome measures:

Table 4.9) when compared to healthy, control animals ($p < 0.0001$, $n = 68$ comparisons). Substantial heterogeneity was observed in the data ($\tau^2 = 891.28$, $I^2 = 98.5\%$). Heterogeneity was not explained by any of the other study characteristic variables investigated with univariate meta-regression, namely: animal, strain and sex of animals, method used to induce animal model, stage of life at measurement and time of assessment.

Table 4.9 Global estimates of effect for the three most widely reported measures of behaviour in developmentally-induced animal models of schizophrenia
Estimates are reported in SMD units, brackets contain 95% confidence intervals of this estimate and N correspond to the number of comparisons contributing to the calculation, nested where appropriate.

	Pre-pulse inhibition	Locomotor activity	Social interaction	Global
Effect size (95% CI)	-0.65 SD (-0.76;-0.54)	-7.97 SD (-16.41;0.47)	-0.61 SD (-0.86;-0.36)	-0.68 SD (-1.00;-0.36)
N	268	68	19	256

4.3.4.2 Treatment-testing studies

Altogether 11 different outcome measures were tested in the treatment-testing studies of developmental models of schizophrenia reviewed here (Figure 4.7). Same outcomes performed in the same group of animals were nested for meta-analysis. The most commonly reported outcomes were PPI (52 comparisons, 955 animals), locomotor activity (20 comparisons, 256 animals) and startle reactivity (18 comparisons, 405 animals).

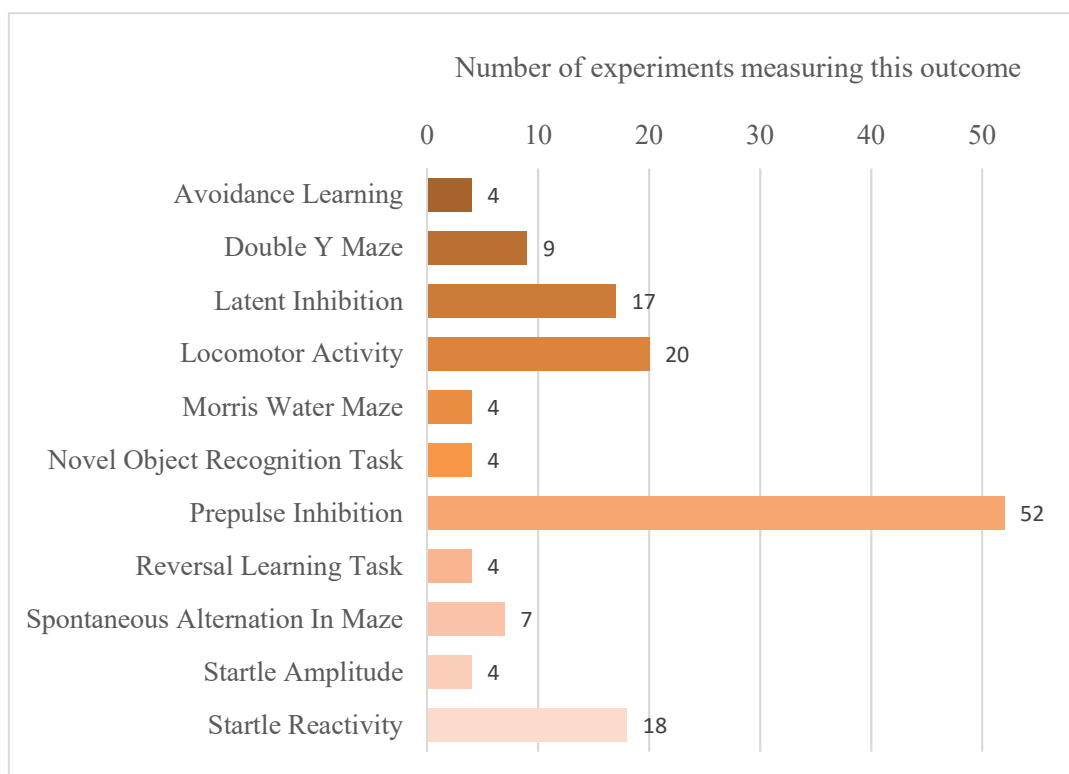


Figure 4.7 Prevalence of behavioural outcome measures reported model treatment-testing experiments of developmentally-induced animal models of schizophrenia

4.3.4.2.1 Pre-pulse inhibition

I identified 52 individual comparisons measuring PPI in treatment-testing studies, using 955 animals. These experiments mainly reported using rats to measure PPI (40 experiments, 643 animals), with 12 comparisons being reported in mice (312 animals).

The rats used were most commonly Wistar rats (37 experiments, 598 animals), with fewer experiments using Sprague Dawley rats (three experiments, 45 animals). The

mice used were all of the C57BL/6 strain, three of which were specified to be of the substrain C57BL/6J (3 experiments, 63 animals). The sex of animals used was mainly male (23 experiments, 361 animals). Much fewer experiments used female animals (14 experiments, 237 animals), or both sexes within the same experiment (15 experiments, 357 animals). The effect of treatment drugs on PPI was predominantly assessed in adulthood (44 experiments, 827 animals), with eight measurements being taken during the juvenile life of the animals tested (128 animals). Background noise used during experiment ranged from 45 Db to 70 Db, with most experiments using background noise at 65 Db (30 experiments, 58% of all experiments). Three experiments did not state the level of background noise used. The pre-pulse used ranged from 2-18 Db above background noise. The main test pulse was always 120 Db.

Overall, treatments improved the performance of animals on the PPI test by 1.33 SD units (95% CI 0.65-2.00) when compared to untreated animals ($p < 0.0001$, $n = 19$ comparisons). Substantial heterogeneity was observed in the data ($\tau^2 = 1.95$, $I^2 = 99.1\%$), but I was not able to explore this any further due to low sample size.

4.3.4.2.2 Locomotor activity

Treatment studies reported measuring locomotor activity in 20 individual comparisons, using 256 animals. Treatments were tested in animal models induced in rats (11 experiments, 136 animals) and mice (9 experiments, 120 animals). The rats used in the experiments were mainly Wistar rats (10 experiments, 114 animals), with one experiment reporting the effect of a treatment in Sprague Dawley rats (22 animals). The mice used were mainly of the C57BL/6 strain (8 experiment, 104 animals), with one experiment measuring locomotor activity in CD-1 mice (16 animals). All treatment assessing studies measuring locomotor activity as an outcome did so in adulthood.

Overall, treatments improved model effects on locomotor activity by 1.40 SD units (95% CI 0.58-2.22), when compared to untreated animals ($p < 0.0001$, $n = 19$ comparisons). Substantial heterogeneity was observed in the data ($\tau^2 = 2.73$, $I^2 = 96.5\%$), but I was not able to explore this any further due to low sample size.

4.3.4.2.3 Startle reactivity

I identified 22 individual comparisons measuring startle reactivity in treatment-testing studies, using 509 animals. These behavioural measures are often measured alongside PPI and assess the animal's fear and anxiety (Hoffman, 2016). For example, it has been reported that increased startle reactivity in rodents might be relevant to early childhood trauma (Jovanovic et al., 2009), which is an environmental risk factor for psychosis (Dean and Murray, 2005).

Experiments reporting this outcome as an effect of treatment did so using animal models created in Wistar rats (10 experiments, 177 animals) or C57BL/6 mice (12 experiments, 332 animals). Startle reactivity was mostly measured in groups of animals that included both male and female animals (12 experiments, 332 animals). Six experiments used exclusively male animals (100 animals) and four used exclusively female animals (76 animals). The effect of treatments was always assessed in adulthood. Startle was measured in response to a pulse at a level of 100 or 120 Db.

Overall, treatments did not seem to substantially improve effects in startle reactivity in animal models of schizophrenia. Overall effects of treatments showed a worsening in behaviour by 0.12 SD units (95% CI -0.70-0.46, compared to other outcome measures: Table 4.10), when compared to untreated animals ($p < 0.0001$, $n = 14$ comparisons). Substantial heterogeneity was observed in the data ($\tau^2 = 0.98$, $I^2 = 98.0\%$), but I was not able to explore this any further due to low sample size.

Table 4.10 Global estimates of effect of behavioural outcome measures in model-characterising experiments
Estimates are reported in SMD units, brackets contain 95% confidence intervals of this estimate and N correspond to the number of comparisons contributing to the calculation, nested where appropriate.

	Pre-pulse inhibition	Locomotor activity	Startle reactivity	Global
Effect size (95% CI)	1.33 SD (0.65;2.00)	1.40 SD (0.58;2.22)	-0.12 SD (-0.70;0.46)	1.01 SD (0.63;1.39)
N	19	19	14	59

4.3.4.3 Outcomes grouped by aspect of behaviour test is intending to measure

Behavioural outcome measures in animal models were also grouped according to human behaviours that these animal behaviours are believed to have relevance to.

4.3.4.3.1 Model-characterising studies

I found that 10 different aspects of human behaviour were measured in model characterising experiments (Figure 4.8). The most common of these behavioural measures measured in developmental models of animals were sensorimotor gating (357 comparisons, 9141 animals), social behaviour (162 comparisons, 2554 animals), anxiety (129 comparisons, 3460 animals), psychomotor agitation (127 comparisons, 2526 animals), and learning and memory (125 comparisons, 2772 animals).

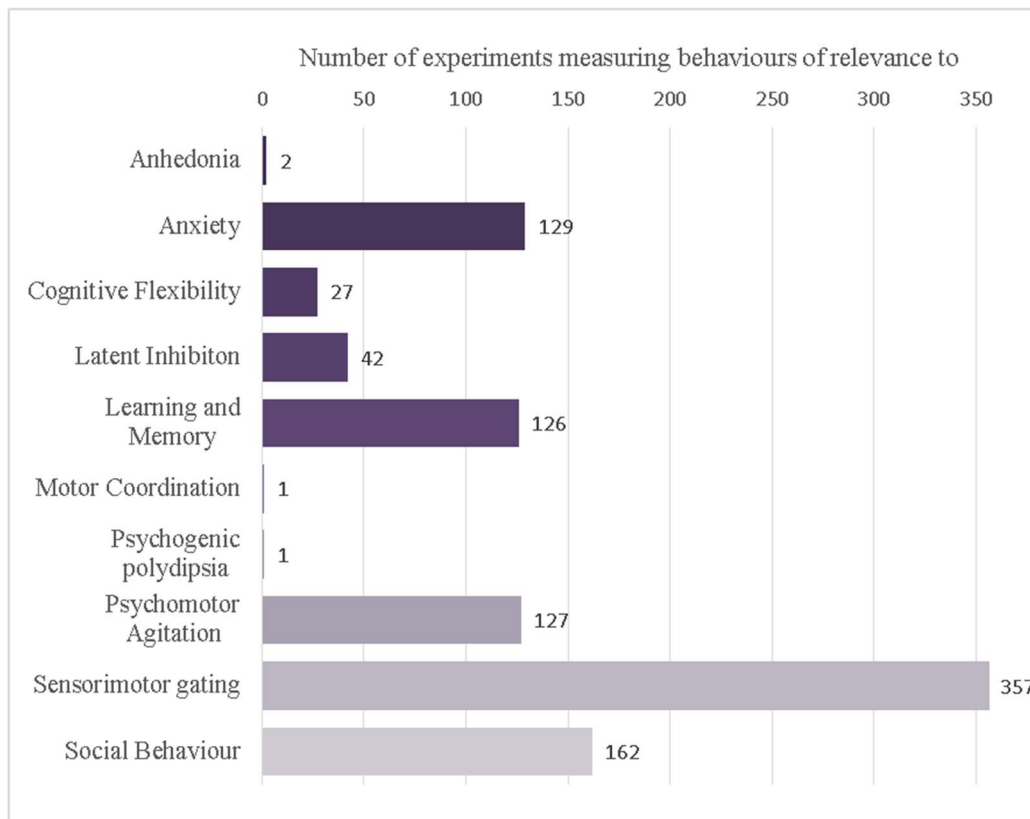


Figure 4.8 Prevalence of human behaviours being modelled through animal outcome measures of behaviour in model-characterising experiments

When data were stratified according to these behaviours the overall effects seen in these measurements in animal models of schizophrenia compared to healthy animals are shown below (Table 4.11).

Table 4.11 Global estimates of effect when stratifying data according to human behaviours being modelled in model-characterising experiments
Estimates are reported in SMD units, brackets contain 95% confidence intervals of this estimate and N correspond to the number of comparisons contributing to the calculation, nested where appropriate

	Sensorimotor gating	Social behaviour	Anxiety	Psychomotor agitation	Learning & memory	Global
Effect size (95% CI)	-0.6415 SD (-0.80;-0.49)	-0.62 SD (-0.83;-0.40)	0.29 SD (-0.43;-0.15)	-6.71 SD (-13.77;0.35)	-0.33 SD (-0.58;-0.08)	-0.68 SD (-1.00;-0.36)
N	90	22	74	78	53	256

4.3.4.3.2 Treatment-testing studies

In experiments looking at the effect of a treatment, I found that six different aspects of human behaviour were measured (Figure 4.9). The most common of these human behaviours were sensorimotor gating (52 comparisons, using 955 animals), learning and memory (28 comparisons, 366 animals), anxiety (22 comparisons, 509 animals) and psychomotor agitation (20 comparisons, 256 animals).

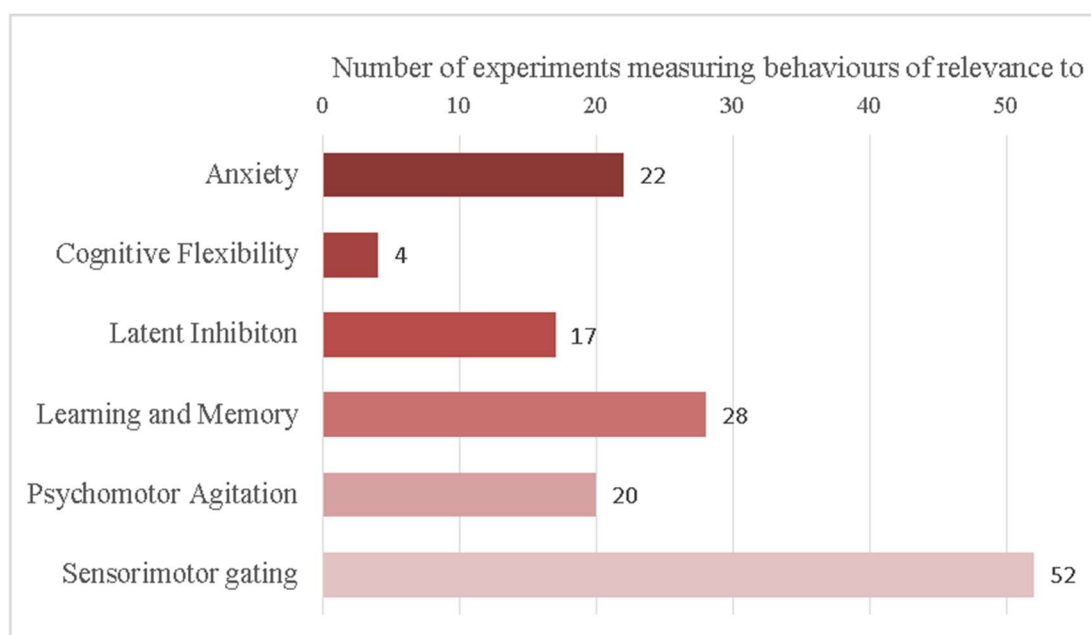


Figure 4.9 Prevalence of human behaviours being modelled through animal outcome measures of behaviour in treatment-testing experiments

When data were stratified according to these behaviours, the overall effects of treatments seen in these measurements in animal models of schizophrenia compared to untreated animals, is shown below (Table 4.12).

Table 4.12 Global estimates of effect when stratifying data according to human behaviours being modelled in treatment-testing experiments
Estimates are reported in SMD units, brackets contain 95% confidence intervals of this estimate and N correspond to the number of comparisons contributing to the calculation, nested where appropriate.

	Sensorimotor gating	Anxiety	Psychomotor agitation	Learning & memory	Global
Effect size (95% CI)	1.33 SD (0.65;2.00)	-0.12 SD (-0.70;0.46)	1.48 SD (0.69;2.26)	0.22 SD (-0.16;0.60)	1.01 SD (0.63;1.39)
N	19	14	20	12	59

4.4 Discussion

Reviewing three widely used developmentally-induced animal models of psychotic disorders, it is clear that these methods of model induction are able to cause an overall negative effect in the behaviours of these animals compared to control animals.

In order to explore their validity further to increase our knowledge about schizophrenia and how to treat it clinically, I looked at how these models were reported, the details of their experimental design, the outcome measures that were measured and the treatments that were reported to have been tested to ameliorate the effects of these models in animals. Moreover, in order to allow for robust translation of knowledge from preclinical to clinical studies we need to measure external validity, which describes the extent to which findings in an animal model can be generalized across populations, environments and species.

I use data extracted in the context of studies describing one or more of three developmental animal models of schizophrenia, to discuss these four levels of validity. Any interpretations made here using these data are done so and should be done so with caution due to often small sample sizes contributing to the overall observed effect.

4.4.1 External validity

In order for pre-clinical studies to be useful for predicting outcome in humans, results in animal studies of psychotic disorders have to be generalizable to humans who present with these psychotic disorders.

Here we see that a large majority of studies use male animals for their animal models. This is a common issue in many pre-clinical research fields (Beery and Zucker, 2011), and can have substantial effects on the generalizability of results gained in the pre-clinical field for other domains of research. While there have been rules in place since 1993 about including women in clinical trials, no such regulations exist in pre-clinical research. Even though some studies report using both male and female animals, results are not always presented and interpreted separately and therefore differences in outcomes are not clear between the sexes. In the clinic, a number of sex differences have been observed among schizophrenic individuals. While males have been shown to have an overall relative risk ratio of 1.6 compared to females, incidence of schizophrenia in males and females varies across different age groups, with incidence of schizophrenia increasing to similar levels in women as in men by late middle age in women (Kleinhaus et al., 2011). Men tend to have an earlier onset of symptoms than women by about 3-4 years (Häfner, 2003), which reinforces the importance of age of model assessment in animals as well as age at which a potential treatment should be tested in these models. Data in this subset of the literature show that while model-characterising studies measure the behaviour of animals at various stages of life, most measurements are still taken in more mature animals and the effects of potential treatments are mainly only tested in adulthood. Schizophrenia has an onset peak in young adulthood, with men generally presenting with symptoms between 18-25, and women usually developing the disorder between 25-35. Women also seem to have a second onset peak after the age of 40, which is likely explained by hormonal changes (Kleinhaus et al., 2011). Moreover, some individuals present with the disorder much earlier or much later outside of this risk period in early adulthood and are potentially different in terms of psychopathology (Sato et al., 2004), something which has not been explored to a great extent in developmentally-induced animal models. Other differences between the sexes that could affect translatability include differences in symptomology. Men and women don't seem to differ significantly on positive symptoms (Ring et al., 1991), but men are observed to have more severe

negative symptoms than women in the form of worse social functioning, blunted affect and avolition. Women on the other hand more commonly show affective symptoms and are observed to respond better to treatments in the clinic (Leung and Chue, 2000). Furthermore, neuroimaging has shown differences between boys and girls in brain maturation during adolescence (Lenroot et al., 2007).

It has also been noted in preclinical studies that males and females of animal models of psychotic disorders respond differently in experiments. For example, females have been shown to show enhanced responses to NMDA antagonists such as MK-801 and phencyclidine in terms of increased locomotion and stereotyped behaviours over their male counterparts. Females also experience less anxiety in response to these drugs compared to male animals (Kokras and Dalla, 2014). Animals also perform differently in some behavioural outcomes, with some evidence showing that female mice show reduced latent inhibition compared to male mice, whereas female rats show greater latent inhibition than male rats (Kokras and Dalla, 2014). While meta-regression did not identify sex of animals as a variable that significantly explains observed heterogeneity in any of the above analyses, focusing research on a single sex is unlikely to give a complete picture of the underlying mechanisms involved in these disorders. As some behavioural measures are sensitive to hormonal fluctuations, these differences should be investigated, and models, outcome measures, and treatments should be validated in both sexes. If we know that there are profound differences between males and females in the clinic, then investigating this variation in detail might increase our understanding of the underlying biological differences. The studies reporting animal models using developmental induction methods reviewed here seem to account for little of this variability in the clinical population of schizophrenia, which potentially affect the external validity of results obtained in these preclinical studies. These variabilities are also seen in different strains and should therefore also be considered when generalizing results gained from mice or rats of the same strain.

Moreover, the experimental design of treatment-assessing studies in animals is often too simplistic compared to its equivalent clinical treatment paradigms. For example, in humans treatments are expected to be taken continuously and discontinuation of treatment is associated with relapse of symptoms (Remington and Kapur, 2010). Simple acute administrations of therapeutic agents in animals do not reflect this treatment paradigm and, as a result, we do not observe the treatment

resistance and side-effects commonly associated with long-term treatment in human schizophrenics.

Findings show that there is large variability in terms of the test used to measure changes in behaviours, the times at which model is induced, and the dose and frequency at which this is done. While variation in study design is widespread across the field and is likely to increase the external validity of findings, many conditions are only reported a handful of times, often within the same study. It is difficult to assess the robustness of findings if results are not replicated across different laboratories. I did not extract information here about certain study characteristics, but as we know the environment is a key player in the development of animals and schizophrenia in humans, it would be interesting to compare differences in the husbandry of animals, especially housing conditions, in models that are primarily based on this factor such as isolation rearing. Of course, this is very much limited by the extent to which these measures are reported within studies and evidence shows that reporting of these variables is incomplete in the literature (Prager et al., 2011).

4.4.2 Construct validity

The construct validity of developmentally-induced animal models of schizophrenia is thought to be high. As mentioned in Chapter 3, one of the supposed advantages of these models is that they are able to recapitulate the clinical picture of schizophrenia being a delayed onset of symptoms, whereby perturbations in early development do not manifest themselves behaviourally till adulthood (Marcotte et al., 2001). While most experiments measured behaviour during the adult phase of animals' lives, measurements taken in young animals showed much larger effects in changes in behaviour between model and control animals. This variable, however, did not explain any of the observed heterogeneity.

While most experiments measured behaviour in animal models created through prenatal disturbances to development, these did not seem to produce as large an effect on behaviour as, for example, infections administered after birth. In humans, prenatal infection has been established as a strong risk factor for schizophrenia, as microbial pathogens have been shown to lead to brain abnormalities, deficits in central nervous system development and associated behavioural disorders in later childhood (Brown and Patterson, 2011). Infections after birth have also been shown to possibly contribute to the aetiology of psychotic disorders like schizophrenia,

although perhaps not as strongly as prenatal infections (Dean and Murray, 2005). Nevertheless, both of these models are likely to be relevant especially as genome-wide association studies (GWAS) show that some genetic vulnerabilities to schizophrenia might be explained as genetic vulnerabilities to infection (Stefansson et al., 2009).

There was a large variety in terms of the experimental design used to induce the model. Animals were prenatally infected at all stages of pregnancy and the time of administration of infectious substances did not significantly explain any of the observed heterogeneity seen in the model. In humans, links between mental disorders and maternal infection have been shown at different stages of pregnancy for different types of infections (Brown and Deskits, 2010). Early to mid-pregnancy is most often implicated (Flinkkilä et al., 2016), however, it remains difficult to tell how this aspect of study design affects construct validity in animal models. Other research suggests that different times of infection in rodents can lead to the appearance of very different results in behaviour. For example, infection at early gestation is linked to sensorimotor deficits in adult offspring, which is thought to be of relevance to clinical positive symptoms, while infection at later stages of gestation is thought to be more closely linked with the development of behaviours related to negative symptoms and cognitive deficits (Macêdo et al., 2012).

Interestingly, the dose of substance administered also did not account for any of the observed heterogeneity in the dataset. The impact of this variable on prenatal infection is also unclear in the clinic. For example, population studies of influenza infections have reported positive correlations between severity of infectious epidemics and number of children born who are later admitted to psychiatric hospitals. However, individual studies do not imply that exposure to an increased dose of prenatal influenza makes it more likely for an individual to then go on to develop the disorder in later life (McGrath et al., 1995). It is clear that in order to fully ascertain the construct validity of an animal model of prenatal infection given at a specific time and of a specific dose, we need to understand more about the clinical impact of these variables.

Adverse rearing seemed to have the least amount of effect in producing a substantial deficit in behaviour in animal models. While adverse child-rearing experiences are considered to be an environmental risk factor for schizophrenia in humans, the correlation between these experiences and actual development of

schizophrenia is not as strong. The relationship becomes stronger when it is combined with a genetic predisposition for the disorder (Schiffman et al., 2001). While I did not review data on anatomical or neurochemical changes in these animals, other studies have shown brain abnormalities in these animals in the hippocampus and prefrontal cortex as well as affecting the microtubular cytoskeleton (Ratajczak et al., 2013). Therefore, this model may still have some construct validity to the human condition. In fact, this model may be more useful over than other models due to the fact that its induction involves no administration of a drug or compound of any kind. Therefore, any treatments that are able to reverse behavioural deficits in these models are likely to not simply be an antagonist to the agonist, or an antibiotic for the specific infection that was used to induce the model in the first place (Geyer and Moghaddam, 2002).

Overall, most experiments described animal models of schizophrenia induced using simple techniques. While the overall effect on behaviour was similar between these experiments and those that induced the model using a combination of approaches, it is arguable that the latter may have more construct validity for the human condition. It is thought that schizophrenia develops as the result of a multi-hit threshold model. Events during a number of key neurodevelopmental milestones give rise to vulnerability factors that individually have a weak effect, but in unison are able to lead to the development and manifestation of symptoms of the disorder in individuals who are already genetically vulnerable (Davis et al., 2016). More studies reporting experimental paradigms of multiple-hit models should be carried out in order to ascertain whether these models might hold more relevance for clinical manifestations of schizophrenia than simple developmentally-induced animal models.

4.4.3 Predictive validity

The number of studies testing a treatment compound in developmental animal models of schizophrenia was substantially less than the number of studies reporting characterising these models. Moreover, most comparisons identified in the literature reported on the effects of already established antipsychotics in clinical use since at least the early 1970s. Many novel or different drugs working via different mechanisms were only reported in small numbers. This was disappointing as it meant I was not able to compare the overall effect of these drugs to the overall effect of the already established drugs.

Clozapine was the most widely tested treatment drug in this subset of the literature, followed by haloperidol and chlorpromazine. Clozapine was also the treatment that showed the largest efficacy in reducing deficits in behaviour when administered to developmentally-induced animal models of schizophrenia. Chlorpromazine showed the second largest effect, although all experiments were carried out in the same laboratory. Haloperidol was the least effective of the three at alleviating behavioural deficits in these animals. In the clinical literature, clozapine is reported to have superior efficacy over other antipsychotics and is especially utilised for the treatment of patients who are resistant to other forms of treatment (McIlwain et al., 2011). Therefore, these results in animal studies seem to be in line with the findings of clinical trials, which show that clozapine performs better than haloperidol and chlorpromazine in chronic schizophrenics (Ravanic et al., 2009), and also better than all other typical antipsychotics (Essali et al., 2009) and atypical antipsychotics (Asenjo Lobos et al., 2010) in general.

In animals, treatments were mostly tested in only male animals. In the clinic, it is not uncommon for men and women to require different treatments at different dosages (Barajas et al., 2015). For example, women require less medication to obtain a similar outcome as men, but they are also more vulnerable to antipsychotic-related side-effects (Smith, 2010). Figures collected in Scotland as part of National Services Scotland for medicines used in mental health show that 54% of patients receiving medication for the treatment of psychoses in 2016/2017 were female and 46% were male (NHS National Services Scotland, 2017). In addition to differential responses to behavioural measures in prenatally infected animal models of schizophrenia documented in the literature, previous studies also show that there are important differences between the sexes neurochemically. For example, it was shown that reductions in glutamate, aspartate and taurine in the prefrontal cortex were more pronounced in male than female animals exposed to Poly I:C infection during gestation (Bitanirwe et al., 2010). These differences between models induced in different sexes can impact the extent to which they can predict the therapeutic efficacy of a treatment and might explain a large number of antipsychotics that prove efficacious in these animal models, but then go on to fail in clinical trials. While the functional mechanisms underlying sex differences remain unclear, what is clear is that in order to increase the predictive validity of any animal model of schizophrenia, treatments need to be tested on animals of both sexes to account for differences, such as fluctuations in hormones (Kokras and Dalla, 2014).

Furthermore, treatments were mostly tested in adulthood. While symptoms of psychosis arise in adolescence and early adulthood, statistics in the literature imply that treatment of these symptoms is often delayed in the clinic. The access of young people to mental health services has been shown to be poor, despite this group of individuals having the highest incidence and prevalence of mental health across the lifespan (McGorry et al., 2013). In fact, data collected in Scotland as part of National Services Scotland shows that the number of patients prescribed antipsychotics is highest for those in middle adulthood (NHS National Services Scotland, 2017). Aside from the issue of ambiguity around how different stages of adulthood in humans actually translate to “adult” animals (Semple et al., 2013), preclinical studies should also focus on testing treatments in younger animals as there has been a lot of interest in the prevention of psychiatric illnesses in the last few years. It is believed that the onset of psychotic symptoms is preceded by psychological or behavioural irregularities. Some researchers believe that the use of preventative treatments may reduce the likelihood of early clinical stages of the disorder progressing to a full-blown psychotic episode (Piontkewitz et al., 2012). In fact, evidence suggests that low doses of second-generation antipsychotics alongside psychosocial treatments may postpone the onset of psychotic symptoms in some individuals (Larson et al., 2010). Furthermore, it has also been shown in the clinical literature that those requiring treatment for the disorder at much earlier (i.e. juvenile and adolescent) and much later (i.e. elderly) stages of their lives are especially vulnerable to showing adverse side-effects in response to antipsychotic treatment (Smith, 2010). This heterogeneity in the clinical population of those requiring antipsychotic treatments, in terms of both age and sex, is not accounted for to a sufficient degree in the preclinical literature reviewed here, making true inferences about the predictive validity of these animal models limited.

Differences in treatment regimens reported in experiments here and the way these drugs are used in the clinic by patients can also affect the predictive validity of the experiments described in the preclinical literature. In the clinic, especially those individuals with more severe symptoms often take antipsychotics long-term (Lally and MacCabe, 2015). Long-term treatment also means that these drugs in the clinic are given using multiple-dosing conditions. From the evidence here, we can see that for the three antipsychotics reviewed, the effect of haloperidol and clozapine was assessed in a large proportion of studies after a single administration of the treatment. It is difficult to assess the long-term efficacy of new drugs this way and it

is not representative of the clinical situation. To further complicate the predictive validity of these animal models of schizophrenia, there can be substantial pharmacokinetic differences between different species as well as strains. For example, the half-life of antipsychotics in rodents is much shorter than in humans, and therefore antipsychotics in animals show very high acute dopamine D₂ receptor occupancy levels and very low D₂ occupancy levels between doses (Kapur et al., 2003). This can be potentially leading to conclusions not representative of the clinical picture even when drugs are administered repeatedly. This is shown by the evidence that a chronic regimen of single injections of haloperidol does not create the same pharmacokinetic profile as the same regimen in humans does (Kapur et al., 2000). Therefore, not only do further studies using developmental animal models need to use experimental designs to test novel therapeutics that more closely resemble the treatment regimens used in the clinic, but they also need to account for neurochemical differences between species. This also introduces the issue of side-effects that are commonly seen in the clinic after prolonged use of an antipsychotic (Leucht et al., 2017). The evidence reviewed here shows that treatment-testing studies generally measure far fewer behavioural endpoints than model-characterising studies and none of these behavioural endpoints included measures that are of relevance to side-effects in schizophrenic individuals. To be able to fully ascertain the predictive validity of an animal model, we have to be aware of both the positive and negative effects of a treatment drug during its development and testing in animals.

Some other drugs tested in the literature included antidepressants, however, I was unable to look at the effect of these due to the small sample of experiments reporting them. Antidepressants have been tested in clinical trials as an adjunctive treatment for schizophrenia aimed at alleviating the negative symptoms of the disorder, however, significant improvement in these symptoms has not been shown (Barnes et al., 2006). More studies would be required both pre-clinically and clinically to assess the utility of these compounds for treatment. Furthermore, no treatments were tested in combined models, which as mentioned before could be a better representation of the clinical picture of schizophrenia. If a treatment drug works in a simple animal model that, has limited relevance to a clinical disorder, it is potentially more likely to fail in later clinical trials. This is a big issue in clinical research as many therapeutic drugs make it to late-stage clinical trials, but don't advance any further (Moore, 2010).

4.4.4 Face validity

In model-characterising studies, behavioural outcomes were measured using a wide variety of different tasks in animals. These tasks were reported to most commonly measure changes in behaviours of relevance to human phenomena and behaviours such as sensorimotor gating and psychomotor agitation, thought to be mainly of relevance to positive symptoms in schizophrenia; social behaviour and anxiety, thought to be of relevance to negative symptoms; and learning and memory, thought to represent cognitive deficits. The face validity of some of these measures of behaviour in animals is of course higher than others, as discussed in Chapter 3. It is not expected that animal models of schizophrenia should try to fully model the entirety of the disorder, especially as in the clinic this is complex and there is a large amount of variability concerning the combination of symptoms experienced by different individuals (Marcotte et al., 2001). Nevertheless, the face validity of those models that are only measured by one domain of behaviour, especially when this is judged based on a single test, is limited for a complex disorder such as schizophrenia, and will certainly not be specific to the disorder. Many of these behavioural endpoints are often measured as part of experiments looking at animal models of other psychiatric disorders. For example, locomotor activity is widely used, but also widely criticised in the preclinical literature for its lack of face validity to clinical behaviours (Powell and Miyakawa, 2006). Increased locomotor activity can also be of relevance to attention deficit disorder or mania, and clinical evidence in fact implies that hyperactivity in the clinic is actually of more relevance to bipolar disorder than schizophrenia (Perry et al., 2010). The lack of specificity of these measures can be a major issue in translation of results as an interpretation of the same measure of increase in locomotion can vary from “hyperactivity” to “agitation” to “increased motivation” based on condition of interest (Moore, 2010).

The most popular outcome measure reported in the literature was pre-pulse inhibition, which measures a phenomenon called sensorimotor gating in humans, which as mentioned previously, is thought to be negatively affected in schizophrenic individuals. As already reviewed briefly in Chapter 3, this has been thought to be a behavioural outcome measure with good face validity as this same phenomenon can be measured in humans using similar stimulus parameters, and response characteristics seem to be similar across different species (Geyer, 2006). Many clinical studies suggest that PPI deficits improve in response to especially atypical

antipsychotic medication in schizophrenic individuals (Leumann et al., 2002), and that clozapine is superior in normalizing PPI compared to other atypical antipsychotics (Oranje et al., 2002). Furthermore, these improvements in PPI have been shown to correlate well with improvements in symptoms, including negative symptoms (Minassian et al., 2007). Face validity of this behavioural measure in animal models is further strengthened by observations here and elsewhere in the literature (Hadar et al., 2015). Data show that deficits in PPI seem to emerge postpubertally, which is an observation also corroborated by clinical evidence (Takahashi et al., 2011), however, human studies in schizophrenia don't imply that this deficit worsens with age (Mena et al., 2016). Furthermore, studies in adolescents at risk for psychosis show that deficits in PPI are present before the onset of psychosis (Quednow et al., 2008; Ziermans et al., 2011), and this behavioural measure in animals might therefore be applicable in the study of preventative approaches to schizophrenia. However, it is also reported that PPI scores of prodromal subjects who go on to develop psychosis are comparable to those subjects who do not (Quednow et al., 2008), although the proportion of those that do is highly dependent on the term of observation. Nevertheless, this reiterates why early treatment of high-risk individuals remains controversial. Furthermore, the face validity of this measure is unfortunately further limited by the fact that a deficit in sensorimotor gating response is not unique to schizophrenia and has been shown to be a neurobiological marker in many other psychiatric conditions including bipolar disorder, Huntington's disease, obsessive-compulsive disorder and attention-deficit disorder (Sánchez-Morla et al., 2016). Moreover, it is also a measure affected by sex as the relationship between impaired PPI and functional affliction is strongest in male schizophrenic individuals in the clinic and therefore is possibly less of a useful predictor of psychopathology in females. While sex did not account for any of the observed heterogeneity in behavioural measurements of PPI here, most experiments did report using male animals. Strain also did not significantly account for any of the observed heterogeneity in behavioural measurements of PPI reviewed here, however, elsewhere in the literature it has been suggested that some strains of rodents exhibit no or only transient PPI deficits after isolation rearing (Geyer et al., 2001). Other factors that should be more closely looked at in future experiments are the role that variables such as pre-pulse intensity, pulse intensity, and background noise can play in variability of outcome. We see from the studies here that these can vary quite substantially between different experiments, and variations in these

variables have been shown in the literature to affect reported outcome (Swerdlow et al., 2000). Nevertheless PPI is considered to be a robust neurophysiologic biomarker for schizophrenia, as agreed by panels of experts, namely MATRICS (Green et al., 2004), and CNTRICS (Carter et al., 2008) (Light and Swerdlow, 2014).

Despite PPI being the most commonly reported outcome in the subset of the literature reviewed here, measuring locomotor activity in animal models of psychosis proved to be a much more sensitive measure of an affected phenotype. This might partly be explained by a few experiments reporting very large differences in this behaviour between model and control animals. Nevertheless, locomotor activity appeared to also be a superior measure of behavioural outcome compared to other measures of behaviour in treatment-testing studies where there were no such outliers in the dataset. Animal hyperactivity as seen in Chapter 3 is a common measure of outcome in animal models of psychotic disorders, despite the fact that it is not a symptom commonly associated with schizophrenia. It is thought that an increase in locomotion in animals in response to novel environments or psychotomimetic drugs is a result of increased dopaminergic activity in mesolimbic and nigrostriatal dopamine pathways (van den Buuse, 2010). Locomotor activity is therefore used to assess for changes in the dopaminergic system in animals and is widely used in the pre-clinical literature as a marker of stereotypical behaviours in the clinic such as psychotic agitation. Where locomotor activity is measured in response to psychotomimetic drugs in developmental animal models, an increased sensitivity to the effects of these agents in these models is thought to model the increased sensitivity of schizophrenic individuals to these same psychotomimetic drugs in the clinic (Powell and Miyakawa, 2006). However, the face validity of this measure of behaviour has recently been questioned as it has been shown that individuals diagnosed with schizophrenia show increased presynaptic dopamine functions in the associative striatum, whereas animal models of amphetamine-induced hyperactivity are mainly driven by dopamine release in the limbic striatum (Kesby et al., 2018). While these models have been shown to be predictive for antipsychotic efficacy, it is argued that this is simply due to a general increase in dopamine function across the striatum when psychotomimetic drugs are administered systematically, and systematic administrations of antipsychotics will therefore have an effect on dopaminergic receptors throughout the brain (Kesby et al., 2018).

More translationally relevant tests for the underlying pathophysiology of psychosis are thought to be cognitive behavioural tasks that can be measured to a similar degree in both humans and animals, and are potentially able to increase our knowledge about associative striatal function in animal models (Kesby et al., 2018). While many studies report measuring locomotor activity or pre-pulse inhibition, far fewer studies report looking at behaviours thought to be of relevance to cognitive deficits or negative symptoms in general. In humans, symptoms associated with deterioration in personal functioning typically arise before psychotic symptoms during a 'prodromal' phase and include deficits in learning and memory, attention, social behaviour, communication and affect (Larson et al., 2010). In fact, neurocognitive deficits at these early stages are thought to be potentially useful in identifying individuals who are at higher risk of developing full-blown schizophrenia (Jahshan et al., 2010). Moreover, some psychiatrists believe that intervention at these early stages is able to prevent psychosis from developing (McGorry, 2015). In humans, cognitive deficits are severe in schizophrenic individuals, averaging about 1-2 standard deviations below the rest of the population (Hurford et al., 2011). The results presented here show that animal models of developmental inductions only show small effects in behaviours measuring learning and memory. Moreover, measures of negative symptoms in the form of anxiety through tests such as startle reactivity and social interaction show small effects in behaviour for both model-characterising and treatment-testing studies. At the moment, current antipsychotics show little impact on these symptoms, and evidence in support of specific treatments for either of negative symptoms or cognitive deficits within the context of schizophrenia is insufficient (Aquila and Citrome, 2015; Remington et al., 2016). It is true that it is often more difficult to model cognitive endophenotypes to the same level of cross-species homology as we are able to do with PPI for example, as in the clinic cognition is often tested through verbal communication (Feifel and Shilling, 2010). Moreover, startle reactivity as a measure of relevance to negative symptoms is also measured differently in animals and humans, which makes the face validity of this measure hard to fully ascertain. Similar to other measures described here, these behaviours are also not specific to schizophrenia. This limits the face validity of animal models for schizophrenia that are only characterized by these deficits in behaviour. Therefore, quantification of models should employ a battery of behavioural tests relevant to schizophrenic behaviours across multiple cognitive and

affective domains to ascertain the full extent of face validity of a model and make results generalizable to subjects in the clinic (Moore, 2010).

4.4.5 Limitations

Unfortunately, as mentioned, resources only allowed for single data extraction. As has been pointed out before, double-data extraction is the gold standard, and therefore some human errors are naturally expected. Data extractors were given continuous guidance throughout the extraction process with any issues discussed with a second investigator (myself). All data extractors followed a pre-defined protocol and I checked 15% of both datasets in detail as the second screener. No substantial and recurring errors were encountered during this process and the data were therefore deemed robust. Unfortunately, due to time constraints, the rest of the data was only 'sense checked' during analysis, where any missing data or any obvious outliers were double-checked and re-extracted if necessary. Clearly, issues like this call for more robust quality control, and advocate for double-data extraction.

While I have reviewed here three widely used and reported developmentally-induced animal models of schizophrenia, the data collected here is not necessarily representative of the rest of the developmental research field and experiments using other methods of model induction. Moreover, data were only extracted for behavioural outcomes and face validity could therefore not be reviewed from a neurological point of view, even though it would be interesting to explore whether these results agreed with data seen in behavioural outcomes about the face validity of these animal models.

We must also remember that conclusions based on the data presented here are affected by the small sample size of experiments for many of the variables measured, which affected my ability to perform multivariable meta-regression and assess for any co-linearity of variables. Finally, non-human animals and humans are different and behavioural changes in one might therefore not necessarily correlate with behavioural changes in the other. Underlying physiological responses to similar insults might not be the same in humans as it is in rodents or even monkeys. And lastly, there are some clear issues over differences in the developmental timeframe of different species, and we do not know the exact impact of this in the context of schizophrenia.

5 Text Mining as a Tool to Aid Systematic Reviews: A New Methodological Approach

This work was performed in collaboration with researchers at the John von Neumann Faculty of Informatics, Óbuda University in Budapest, Hungary. As the development of this work was a computational challenge, rather than a biomedical hypothesis testing experiment, the reporting of the work will not be in the style of traditional biomedical reporting. I have instead based the structure of this chapter on that specified for dissertation projects at the School of Informatics of the University of Edinburgh (<http://www.inf.ed.ac.uk/teaching/courses/diss/guide.html>).

5.1 Introduction

As a result of advances in technology, the speed of inception of primary evidence is bypassing the speed at which we are able to summarize this data. This is a real limiting factor when it comes to making evidence-based decisions in medical biology, a problem that has been highlighted in the field of clinical trials (Bastian et al., 2010), but is an equally big issue in animal studies.

The process of manual data extraction and analysis, is in itself a slow and subjective process, but these issues are magnified when it comes to the manual analysis of large volumes of data (Fayyad et al., 1996). This is especially the case in larger systematic reviews, which aim to incorporate all of the literature available for a given research topic and take as much information from these sources as possible. This contributes to the increase in the amount of data needing to be processed: 1) the number of records needing to be reviewed (i.e. studies in the review) and 2) the number of fields of interest that we are attributing to each of these records (i.e. study characteristics) (Fayyad et al., 1996). As a result, the benefit of systematic reviews is diminished by the time taken to perform them, so that findings of these reviews are generally out of date by the time of publishing. Adding to this time lag, is the concept that in systematic reviews it is encouraged that studies are reviewed by two independent investigators to account for human error and increase reliability.

5.1.1 Motivation behind work

I found reviewing and annotating publications a time-consuming task for the number of studies included within my review. This was an underlying issue throughout my

project, I realised that large, and broad projects would always have this limitation, unless alternative approaches were considered to enhance the speed of the review process. I hypothesised that automating part of the review process might make a difference to the speed at which you could complete a specific stage of the process. Such a computer aided approach would not only help a reviewer make more efficient use of their time, but also in turn maximise the potential of results gained from these reviews and let available evidence inform subsequent stages of the clinical drug development process faster.

5.1.2 Objective of work undertaken

My aim was to design a program that would not only reduce the aspect of reading and analysing publications for reporting of risk of bias items, but also reduce the impact of naturally occurring human error when doing this for a large number of studies.

My two main objectives in order to achieve this: 1) develop a set of term identifiers that would allow for automatic phrase recognition, and 2) create a program that I could easily use to assess these words within multiple files at once and update a database storing data on these files accordingly. Primarily, I wanted to find a solution for increasing the efficiency of my own work, but if successful, also produce a tool, which could be of use to other systematic reviewers faced with similar problems when processing a large number of publications.

This project was a subsidiary investigation to my overall systematic review of the field, but in this chapter, I describe my reason for developing this methodology, discuss the utility of the approach taken and the potential of incorporating computer-based approaches such as this one in the field of systematic reviews in pre-clinical biomedical sciences. While the key aspect of this work was the actual success of predicting risk of bias items within a number of large pieces of text reliably, I also describe the design of the project that we created to allow us to test this.

5.2 Background

Fayyad *et al.* (1996) suggested some time ago that volumes of this magnitude of work should be automated at least partially, instead of humans carrying out this process. They introduced the idea of knowledge discovery in databases to tackle

this problem of increasing difficulty in digesting the extensive volume of information that the digital age has made available (Fayyad et al., 1996).

This involves the use of various computational tools to aid the process of humans acquiring useful knowledge from large volumes of digital data. A major step in this process is the use of data mining (Fayyad et al., 1996). This is the use of computational algorithms for the identification and extraction of patterns from large pieces of data, which are then often used for interpretation of the data and in aid of decision-making and derivation of useful knowledge from these data (Fayyad et al., 1996; Hearst, 1999).

A potential area where data mining can be used is in the mining of free-form text (i.e. text that is unstructured, for example those in a word processor). This is a form of natural language processing (NLP), where it is possible to automate annotation and thus classification of this text (Fayyad et al., 1996) by discovering information within data through the establishment of patterns across datasets and separating out signal from noise (Hearst, 1999). This is precisely what we do when we analyse a publication, we try to filter the relevant information from the background noise.

Most NLP applications to systematic reviews seem to focus on the screening and inclusion/exclusion of relevant articles into a dataset. While there has been a number of attempts to utilise natural language processing techniques in the automating of data extraction in systematic reviews, these have been mainly exploratory, and few have been so far utilised in actual reviews for a full or even partial automation of the data extraction process. Moreover, the extent to the number of potentially extractable data elements that have been investigated is far from complete, and notably few automated natural language processing techniques focus on automating the assessment of the internal validity of publications (Jonnalagadda et al., 2015).

5.2.1 Work carried out so far and possibilities for improvements

In light of the size of my dataset and the need for a second screener during phase 2 of the project, the motivation for this project was to develop a more efficient tool to aid my work by reducing the time taken to review full-text publications and provide a secondary form of review.

Based on this, I recognised that annotation and categorization was a stage of the systematic review process where text mining could potentially contribute. More specifically, I knew that one of the most uniform variables that could be extracted from a publication is the evidence for the reporting of measures taken to reduce the risk of bias and other methodological criteria. This was mainly based on personal experience of searching documents for specific words while usually performing this stage of the review. As there are generally only a limited number of ways in which certain risk of bias and methodological quality criteria can be described, by creating a list of search words we could search PDF's of publications to help classify whether they had reported or not reported some of these list items.

A solution that standardized and generalized the approach to this process would have the opportunity to overcome and minimise these issues within this seemingly simple stage of the review.

5.2.1.1 Current method of reporting of risk of bias assessment

The instruments used to assess the reporting of internal validity can vary considerably in the field of pre-clinical research (Krauth et al., 2013). Fundamentally though, they all use the same basic approach to assessment: scoring for the presence or absence of the clear description of a risk of bias or study quality item within a publication (Horn et al., 2001). As mentioned before, I use a modified CAMARADES checklist in my own review (Macleod et al., 2004).

Once key words are identified within the active document of focus, I read the surrounding word environment and make a judgment call about whether there is sufficient evidence for the correct reporting of measures taken to reduce each risk of bias item and other methodological criteria on my checklist. I do this at a full-text level and while these measures are usually described within the Methods or Results, I am still required to find the right phrase or lack of thereof within each piece of text.

As the analysis of this data element is a simple 'Yes' or 'No' interpretation of whether each item of interest has been reported or not, this should be a stage of the review where annotations should be consistent across different investigators. In practice, discrepancies often arise even at this stage of the review process. It is possible for a reviewer to miss the key piece of text required to interpret reporting of criteria. Secondly, subjectivity is introduced during interpretation of information.

For example, methods of randomisation are rarely described (van der Worp et al., 2010), therefore a reviewer has to take reports of these items at face value and judge whether wording explains a generally correct method of randomisation or not. Finally, repetition of the process for large number of publications increases the chances of human error arising and the impact of subjectivity.

By pre-defining which expressions should be classed as valid examples of the reporting of items on my checklist and which ones should not, a tool could potentially be created which automates this stage of the review. A solution of this kind would take out the above-mentioned subjectivity of text interpretation, and reduce the impact of the small percentage of human error that inevitably occurs when processing large amounts of data. A tool like this would have the potential to eliminate the need for a second screener, which is a common method used to try and reduce the likelihood of introducing errors into the dataset (Barchard and Pace, 2011). In addition, the tool could be used as a way of validating the work of a human. This was a key requirement for my project, as I was not able to have a second data extractor for this part of my project.

5.2.2 Previous work in similar fields

Automation of risk of bias item reporting assessment in data identification and extraction from clinical trials has been explored so far by two different research groups (Marshall et al., 2015; Millard et al., 2016). Marshall *et al.* (2015) for example used machine learning to automatically judge the risk of bias within a trial in the COCHRANE Database of Systematic Reviews as well as extract text fragments supporting these conclusions (Marshall et al., 2015). This approach was similar to those used by researchers at the University of Bristol, who used machine learning to find relevant sentences within a piece of text and use this to rank both sentences and articles according to risk of bias (Millard et al., 2016). Both groups used supervised machine learning, with mean precision scores ranging from 0.53 to 0.87 for items in both studies combined (Marshall et al., 2015; Millard et al., 2016). Unfortunately, both of these solutions have only been developed for the assessment of human trials. Both mainly focus on randomisation, allocation concealment and blinding at full-text level (although Marshall *et al.* also include inclusion/exclusion reporting), which are all risk of bias elements that are transferable to the pre-clinical research field. Despite this, some key aspects in the reporting of these risk of bias items differ between clinical trials and animal studies (Krauth et al., 2013). For this

reason, I wanted to see if I could develop a similar tool to automate the assessment of the reporting of risk of bias for studies reporting pre-clinical studies as opposed to clinical trials.

Firstly, the approach taken for the evaluation of measures taken to reduce risk of bias is different for clinical trials, and the one that many systematic reviews use when assessing risk of bias in animal experimental publications. The approach for risk of bias prediction for clinical trials, and incidentally the tool used by both groups to develop their algorithm, is based on the COCHRANE Risk of Bias tool (Higgins et al., 2011). This tool does not class articles based on a simple checklist, but rather gives studies a “high”, “low”, or “unclear” risk of bias based on the amount of detail given within a published trial or mixture of sources for the same trial (Higgins et al., 2011). In comparison, many systematic reviews of pre-clinical studies, including this current work, have used checklists to assess the likelihood of risk of bias or low methodology (Macleod et al., 2004; Sena et al., 2007). Methodology is evolving and newer tools, such as SYRCLE’s risk of bias tool (SYRCLE’s RoB tool) for animal studies, which have been developed based on the Cochrane RoB tool, offer a more consistent and similar approach to the assessment of risk of bias to those used for human studies (Hooijmans et al., 2014). As it makes an important differentiation between assessment and interpretation, it is possible to class evidence of not randomising separately from those that did not give enough information to determine whether randomisation took place or not. This is not something current checklists, like the one used here, usually account for. While the applicability of this tool will still depend on the current detail of reporting within pre-clinical studies, moving towards this level of reporting of risk of bias assessment has the potential to yield a more accurate critical appraisal of the methodological quality of animal studies than a simple checklist method. While the content of description used still matters within these animal studies when it comes to interpretation of risk of bias within a study (Kilkenny et al., 2009), a number of challenges remain in analysing risk of bias in animal studies and why we analyse clinical trials differently to animal studies.

Most importantly, reporting of risk of bias has improved in clinical trials (Plint et al., 2006), while there has been little improvement seen in animal studies (Macleod et al., 2015). This is possibly due to more advanced awareness around the topic of methodological quality in clinical trial designs, and how inadequate studies can erroneously estimate treatment effects (Schulz et al., 1995); as well as the

introduction of guidelines adopted by journals (Schulz et al., 2010). The literature on how to evaluate risk of bias in a randomized clinical trial is also more extensive, whereas there was little in terms of guidance for the animal research community before the introduction of the ARRIVE guidelines in 2010 (Kilkenny et al., 2010b).

This lack of universal standards and concurrent unfamiliarity with the rationale of these certain aspects of experimental design (Festing, 2013), mean that pre-clinical evidence reviewers come across a wide range of different expressions used to report risk of bias items, where sometimes these explanations don't actually qualify for the true definition of the items (Hooijmans et al., 2014). Moreover, in animal studies we have greater flexibility in terms of experimental design including multiple species and models to recapitulate human behaviours, as well as a large range of outcomes measured. This is often due to animal experiments being exploratory (Ioannidis, 2012). While the COCHRANE risk of bias tool, encourages the assessment of risk of bias for each main outcome or class of outcomes (Higgins et al., 2011), many systematic reviews of animal studies perform reporting of risk of bias assessment on a publication level (Sena et al., 2007). While some forms of assessment suggest doing this for individual outcomes within studies (Hooijmans et al., 2014), this can sometimes be difficult as these risk of bias items are rarely stated separately for different outcome measures. These are all differences that can make the automation of pre-clinical studies compared to clinical studies different, and perhaps more difficult.

In addition, in a systematic review of pre-clinical studies, we disregard human data if there is any within the same publication. This is something that needs to be taken into account when developing a tool for the identification of specific terms within a publication, so as to try and not pick up on risk of bias reporting for human studies within the same publication.

5.2.3 The relevance of this work to our systematic review process

CAMARADES has been working on creating a fully integrated online platform, which aims to act as a hub for systematic reviewers of preclinical studies to carry out different stages of their review using practical web-based tools and applications (www.syrf.org.uk). I wanted to create a solution that could potentially be further developed and incorporated into this framework if proven successful. This would add an additional step to the already present online systematic review process that

includes a web-assisted tool for screening and annotating your dataset. This approach of applying text mining to reporting of risk of bias assessment is in line with work by the SLIM (Systematic Living Information Machine) consortium, a multi-centre collaborative pilot project that aims to automate various stages of the systematic review process with the motivation that someday automatic “living” systematic reviews will be achievable.

5.3 Work undertaken

5.3.1 Approach to design

The main concept of the work I undertook was to use natural language processing to review a collection of publications describing pre-clinical studies of biomedical disorders, by identifying and predicting risk of bias assessment for each study.

There were two components to the solution designed: the search terms used to capture the relevant information within a piece of text, and the program that ran these terms in a given set of files.

5.3.1.1 Objective 1. Creation of a set of term identifiers

The application of natural language processing to data extraction in systematic reviews is dependent on term identification, which is the main limiting factor in the processing of useful information in the literature (Krauthammer and Nenadic, 2004). My first objective was to develop a set of expressions that would allow for automatic term recognition within a text. There are a number of different approaches to this (Ananiadou and McNaught, 2006), including the ones used by the studies mentioned above where the machine learning approach has been used to identify relevant sentences within a large piece of text (Marshall et al., 2015; Millard et al., 2016). Because our interpretation of a study’s risk of bias in animal studies requires a simpler approach of assessing for a positive or negative reporting of a checklist item, we took two different approaches and compared the results of these for their ability to recognise terms within a large piece of text.

The first approach to term recognition, the dictionary-based approach, works by using existing terminology in order to find the occurrence of a term (Ananiadou and McNaught, 2006). The system is given a list of terms (i.e. could be a string) and if that term is matched within the text of interest, then it is recognised by the computer

as an instance of that term and its related concept (Caporaso, 2009). The accuracy of this approach to term recognition is most affected by term ambiguity, and deviation from the original terms specified (Ananiadou and McNaught, 2006). In comparison, the rule-based approach uses patterns of term formation, where rules are used to characterize a common structure for terms (Ananiadou and McNaught, 2006). Here a string of defined characters is matched within a text of interest against a rule or rules that are given to the system. This is often achieved by way of regular expressions (Caporaso, 2009).

I developed a set of key phrases for three widely applicable approaches, which can reduce the risk of bias in experimental design and provide more precise results. These were random allocation to treatment group, blinded assessment of outcome, and sample size calculation (SSC). I chose these risk of bias and methodological quality items because they were all included in the minimum core set of items authors should report in their grant applications and scientific publications (Landis et al., 2012).

As I did not have structured data (i.e. quotations taken from actual publications) available for randomisation, blinding and SSC (i.e. specific quotes within the publication to justify risk of bias assessment), I had to first distinguish useful terms of the subject from non-related noise. From previous experience working on other systematic reviews, I realised that simple search words would not be specific enough for the computer to be able to always make a correct prediction for an item. For this reason, I decided to assess the phrase environment of words commonly associated with my items of interest. This was then used to develop a set of key phrases that would be designed to predict reporting.

For randomisation and blinding, I used the library of PDFs I had available from my systematic search of animal models of psychotic disorders as my training set. I initially made a list of keywords that are commonly used to describe each of these items. The idea was to use the program PDF Xchange Viewer, find phrases containing those keywords and score them as “yes” or “no” for the reporting of that study quality item. I used “random” for randomisation and the terms “blind”, “aware” and “naïve” for blinded assessment of outcome. I made a record of every word environment I had identified these terms in in my set of .pdf files and used it to create a list of phrases that I would always class in a paper as having reported that risk of bias or methodological quality item in question. I used the definition of

randomisation as a random allocation of animals used to control and intervention groups. Blinding was defined as the reporting of investigators being blinded to the knowledge of which intervention had been given to which group of animals during the experiment (Hooijmans et al., 2014). This means that a sentence that said “animals were tested in a random order”, I would not class as having randomised during experimental design, but a sentence such as “animals were randomly assigned to groups” I would class as an example of reporting of randomisation. Similarly for blinded assessment of outcome “whole-cell blind patch clamp technique” should not be an example of reporting of blinding, however, “tests were carried out by a blind investigator” should be classed as a true example of the reporting of blinding, in my opinion.

Due to the fact that SSC is a methodological quality criterion that is seldom reported in pre-clinical research fields (van der Worp et al., 2010), my approach for search term development was slightly different here. I used data extracted in our CAMARADES database as part of other systematic reviews to identify publications that had been classed as reporting SSC. Publications included here described pre-clinical models of stroke, Parkinson’s disease and multiple sclerosis. While I relied on other investigator’s classification of this methodological criterion within a study, when identifying relevant publications, I made my own interpretations of each publication I reviewed. Much like before, I made a list of expressions that were used by authors within publications to say they had performed SSC.

I extracted the shortest possible textual span that led me to determine reporting of risk of bias, even if these were not full sentences. It was essential that I obtained as many examples of descriptions as possible, so I extracted all possible standalone phrases that led me to my conclusion, even if this meant extracting multiple instances from the same publication.

5.3.1.1.1 Dictionary-based approach

Using my final list of differentiated word environments, I narrowed these down to a list of phrases that would pick up as many of these examples as possible while trying to avoid picking up any of the examples where risk of bias items would not be scored as having been reported. These phrases were then used as our term recognition set to parse files with.

5.3.1.1.2 Rule-based approach

Regular expressions are a sequence of symbols and characters that declare a string or a pattern to search for in a longer piece of text. I thought that the use of regular expressions would provide more flexibility over simple phrase searching in terms of picking up on expressions that vary slightly from the list of expressions specified. It would also exclude instances where these expressions are used to explain something that was not done (i.e. experiment A was not randomised). I created a list of regular expressions that would serve the same function as my original term recognition set of words above, but without the limitation of having words appear in the specific order, case and tense given.

5.3.1.2 Objective 2. Creation of program to run and record results of term recognition

My second objective was to design an application that was able to perform and automate a series of steps that mimicked the human reviewing process (Figure 5.1) for multiple studies at once. This could then “read” a PDF, using our developed set of term recognizers, identify any potential useful snippets of text, predict risk of bias items of interest for that study, and finally record this information within a database containing data for a multitude of studies.

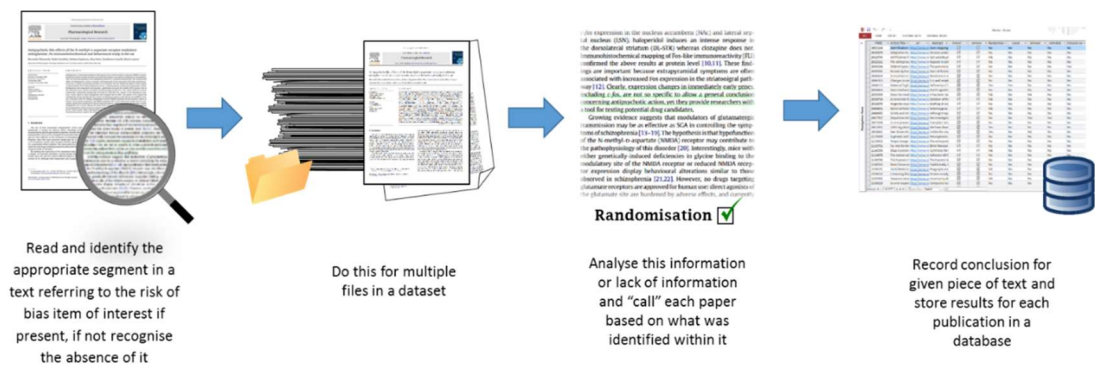


Figure 5.1 Key stages of applying text mining to the process of reported study quality assessment

To achieve this, we needed the program to have the following three functionalities: (1) read a given file; (2) search for words or phrases in this file; (3) interpret results of search and record this in a database accordingly.

5.3.1.2.1 Specification

5.3.1.2.1.1 Usability

Primarily, I was the user, but it is a possibility that systematic reviewers in other fields of pre-clinical research could use the solution one day. I had to take into account certain basic characteristics about these potential users: variability in familiarity and ease of working with computer applications, varying levels of knowledge and experience of systematic reviews and reporting of risk of bias items, as well as different locations of users and resources including computer platforms preferred and available to them.

5.3.1.2.1.2 Input and Output

While it has been stated that there are numerous instruments available to assess risk of bias in pre-clinical studies (Krauth et al., 2013), systematic reviews performed by the team at CAMARADES and collaborators, tend to centralize around existing, previously published study quality lists of assessment (Macleod et al., 2004; Sena et al., 2007). These are then often adapted to the research area of interest studies that review is being conducted in and can assess a variable number of different risk of bias items (Krauth et al., 2013). Therefore, this part of the application would be different for different investigators and I wanted to make the input of this information open to different sorts of options.

In terms of the information output that would be beneficial to the user, the aim of this application was primarily to show the search results for each individual publication (i.e. phrase identified and final 'call'). This was so I could assess the overall effectiveness of our solution, but also in order to further develop the program by knowing which phrases had been identified within each publication and be able to look at those in more detail where disagreements arose between the computer and the human screener. In contrast, other users might only be interested in a summary of the prevalence of a certain risk of bias item, therefore, this sort of output was also important. I wanted the output to be communicated to the user through the user interface and give the user the option of storing the results in a database. As my project uses MS Access to record information about studies within my dataset, I wanted to create a solution that would be able to update this sort of database primarily.

5.3.1.2.1.3 Adaptability

CAMARADES has been working on creating a fully integrated online platform as mentioned earlier; therefore, I wanted to create a solution that could be incorporated into this framework if needed in the future.

5.3.1.2.1.4 Summary of software requirements

Based on these criteria with the help of a programmer I created an application that would ideally meet the following requirements to be considered successful:

- User-friendly interface
- Allow the investigation of the reporting of risk of bias item of choice
- Search a number of different files at once
- Facilitate the storage of these results in a database, if necessary
- Possibility of integration into existing web-based platform

5.3.1.3 External software

One of the biggest considerations before starting the development of the project, involved thinking about how articles in a systematic search dataset might be “read” by the program. Usually scientific publications are published in PDF (Portable Document Format) (Adobe Systems Incorporated, 2006). This file format can present elements of a printed document as an electronic image that is independent of software, hardware and operating system. PDF is based on PostScript, a page description language that runs in an interpreter to generate an image. This image is coded using separate objects, meaning text and images are stored in a special format that is fixed and not easily modifiable. This creates an obstacle when it comes to analysing these files, as the text that we see is not encoded in a plain text format. For this reason, in order to be able to edit or use this text of interest, it needs to be converted to an alternative format before it could be analysed.

PDF documents were converted to plain text format using an external software. We trialled multiple converters available on the web and chose the “best option” (<https://bytescout.com/>). This software was chosen as it was able to convert as many of the given PDF files as possible (including secured PDFs in some cases) and gave an output that was meaningful and therefore could be easily searched using whole phrases (i.e. read paragraphs into file in the right order, with no breaks in-between).

5.3.1.4 Proposed Validation of Solution

I tested the effectiveness of the program and validated my search terms developed for risk of bias items by comparing it to previous scoring of risk of bias by a human investigator (i.e. the current “gold standard”). For this, I used a separate dataset of 1359 of other *in vivo* studies from another project carried out by our group (Macleod et al., 2015), where risks of bias had been previously ascertained by two independent human reviewers. Results of the human screening process were stored within an independent MS Access database, where the application could update the appropriate fields based on the results of the search. Once the results of the application were known, I could simply run a query in order to look at the number of studies in agreement and those where the computer did not give the right prediction for risk of bias.

In order to evaluate the performance of the search terms developed I compared the predictions made by the computer to the values given by the human screeners and looked at agreements and disagreements through a simple confusion matrix (Figure 5.2).

		Actual values (Human screener)	
		Yes	No
Predicted values (Computer classification)	Yes	True Positive	False Positive
	No	False Negative	True Negative

Figure 5.2 Confusion matrix showing how performance of computer tool is evaluated against human screener

Using values from this table, I was able to calculate sensitivity (i.e. also termed recall). This is the number of positive publications correctly predicted by a computer, as a proportion of all the positively classed publications by a human screener (Equation 5.1);

5.1

$$\frac{\textit{True Positive}}{\textit{True Positive} + \textit{False Negative}}$$

Alongside specificity, which is the number of correctly predicted negative publications by a computer, as a proportion of all the actually negatively classed publications by a human screener (Equation 5.2);

5.2

$$\frac{\textit{True Negative}}{\textit{True Negative} + \textit{False Positive}}$$

And Precision, which is the number of correctly predicted positive publications by a computer, as a proportion of all the positive predictions it has made (Equation 5.3).

5.3

$$\frac{\textit{True Positive}}{\textit{True Positive} + \textit{False Positive}}$$

Finally, I calculated an F-score, which is a harmonic mean between sensitivity and precision, and thus gives equal weight to both. As a result, it requires both values to be high in order to produce a high value (Equation 5.4).

5.4

$$2 \times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}}$$

Altogether, these performance measures can give us an idea of how good the tool is at correctly predicting positive instances of a term and thus concept (i.e. sensitivity) and the effectiveness of the tool in correctly not identifying those instances where no concept is to be associated (i.e. specificity). Finally, it gives an idea of how sure we can be that when a computer makes a positive call, it is a correct prediction (i.e. precision) (Sokolova and Lapalme, 2009).

5.3.2 Design of Solution

Here, I describe the overall structure and functional flow of data within the solution. The exact code can be found within the appendix of this document including annotations of the function of each line of the code (see *Appendix III. Text mining tool: code used for program and list of expressions*).

5.3.2.1 Design features

5.3.2.1.1 Platform solution

The system was designed in a globalised manner through a web-based solution, with the thought that it could be incorporated into and hosted on existing web platforms. Here, once public access is given to the web page on a local or a global network, multiple users could use the application simultaneously. The advantage of this approach over a localised solution would be that data would be stored centrally on a server and individual user access would be controlled. This also allows for more frequent back-ups to be made centrally and thus keeping data safer, even when multiple people are working on the same project.

5.3.2.1.2 Architecture of the system

The architecture of the system was built in the architectural style of a Representational State Transfer Application Programming Interface (REST API). This is widely used for designing dynamic data driven web solutions. This technology allows the execution of required tasks within the program to be distributed (i.e. split up) between a front-end and a back-end. The back-end of the program performs all the core functions that can be automated such as tasks that require large calculations. The front-end takes input from the user that specifies instructions on how to execute back-end tasks and displays results of actions on a user-friendly interface. This sort of separation of tasks helps to improve performance of the tool. It also allows different browsers to be able to display a standardized user interface on every platform (Figure 5.3).

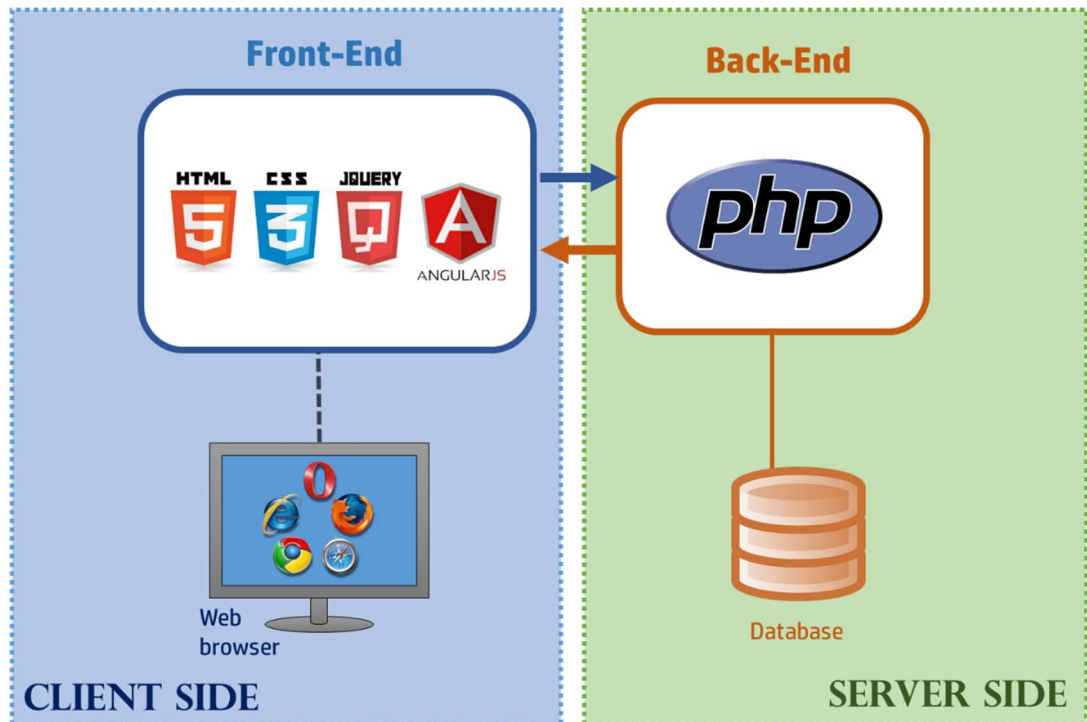


Figure 5.3 The architecture of the program based on a REST API design that allows for complicated tasks to be separated out and performed “behind the scenes” by the back-end, while communicating only the relevant data with the front-end

The two ends communicate with each other through standardized message structures (i.e. written in JSON, described in more detail below), which are sent back and forth between the two ends. These are in the form of GET and POST data, forms of AJAX HTTP requests (Asynchronous JavaScript and XML Hypertext Transfer Protocol), which are a set of rules used to transfer data between clients and servers. For example, GET is used to request data and POST is used to submit data. When the front-end instructs the back-end to do something it will prepare instructional statements in a standardized format that sends data to the back-end via a POST message. This way, the back-end knows what data to work with and it can carry out the appropriate calculations and actions. Once the back-end has carried out these actions, it then sends a performance report back to the front-end to say whether it has successfully carried out the specified instructions or if any errors occurred.

5.3.2.2 **Functional features**

5.3.2.2.1 Key stages of system behaviour

The application is able to carry out three functionalities:

1. Perform a search of a given set of .txt files using simple phrase expression, regular-expression or both
2. Check size of given set of .txt files and compare these files with another set of .pdf files to highlight those .txt files of small size* and those which do not have a corresponding .pdf file

Update a database (i.e. MS Access) using results obtained from the previous two steps.

**File of a small size was defined as anything below 1KB. This equates to 1024 bytes, which roughly equates to the same number of characters. One page of text is equal to about 3290 characters. This limit of file size was decided on from the observation that when comparing files before and after conversion from .pdf to .txt files, .txt files smaller than this were generally not converted properly and therefore did not contain any relevant and usable information for the text mining to be validated on.*

The program executes each of these actions so that it follows the overall communication system specified above. When the user prompts the program to do something:

1. Front-end prepares an Update Statement containing instructions that is submitted to the back-end (i.e. search string to be processed and ran on files)
2. Back-end receives this statement, it executes the statement accordingly, by looping through all relevant files (i.e. actual running of search)
3. Back-end returns results of action to front-end, so it can be displayed to the client (i.e. results of search displayed to the user).

This setup allows data to theoretically flow through the program to create a user-friendly tool that is able to search publications using search expressions provided by the user and display results of this search back to the user, allowing them to update this data to a database (Figure 5.4).

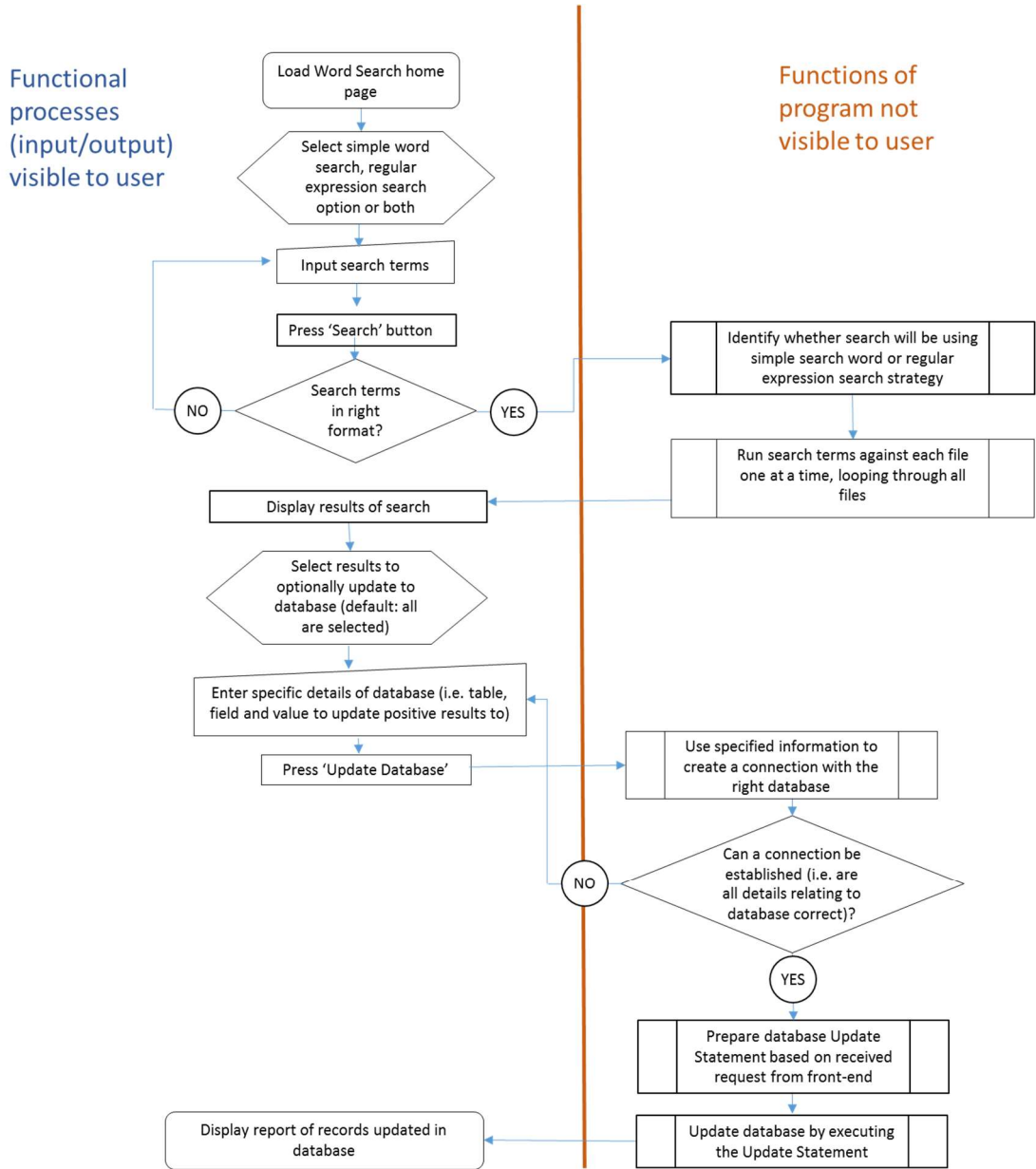


Figure 5.4 Detailed flowchart of how the search function of the tool works – including how front-end and back-end communicate

The file checker was a later addition to the application and functions in a similar flow, however, is independent of the search (Figure 5.5).

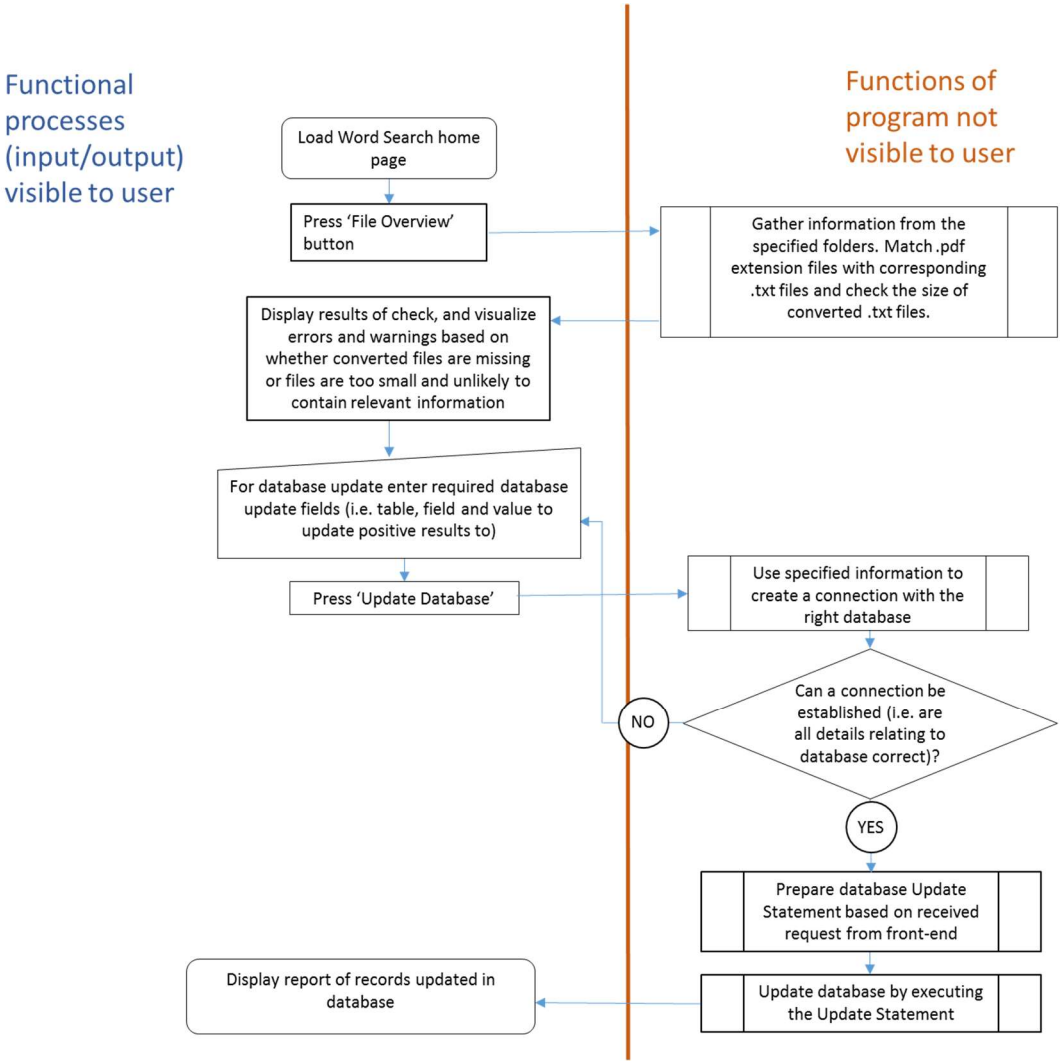


Figure 5.5 Detailed flowchart of how the file check function of the tool works – note that the database update is the same as for the search function

5.3.2.2.2 Graphical appearance and functions

The application is achieved through a single page application. This means that on initial loading of the page, the whole application is loaded into the browser. Due to the REST API architecture, however, only certain information is made visible to the user at any given stage of application use. This is controlled by the front-end process workflow, describing which components should be made visible to the user and which should be hidden from the user interface at different stages.

5.3.2.2.1 Welcome page

On loading the application the first 'page' that is visible is the Welcome Page of the tool (Figure 5.6). From here, the user has the following interactive possibilities:

1. Perform a simple phrase expression search – this includes a free-text box where the user is able to enter their search phrases separated by a comma; and the actual "Search" button*
2. Perform a regular expression-based search – this includes a free-text box where the user is able to enter their regular expressions in a specified format; and the same "Search" button as before*
3. Check existence and size of files available in the specified folder – this function exists so user is able to check whether converted .txt files exist for each .pdf file in a given folder and whether these .txt files are of a size that are likely to contain valuable information

**There is an option to execute both searches, by selecting the tick-box and entering data for both options.*

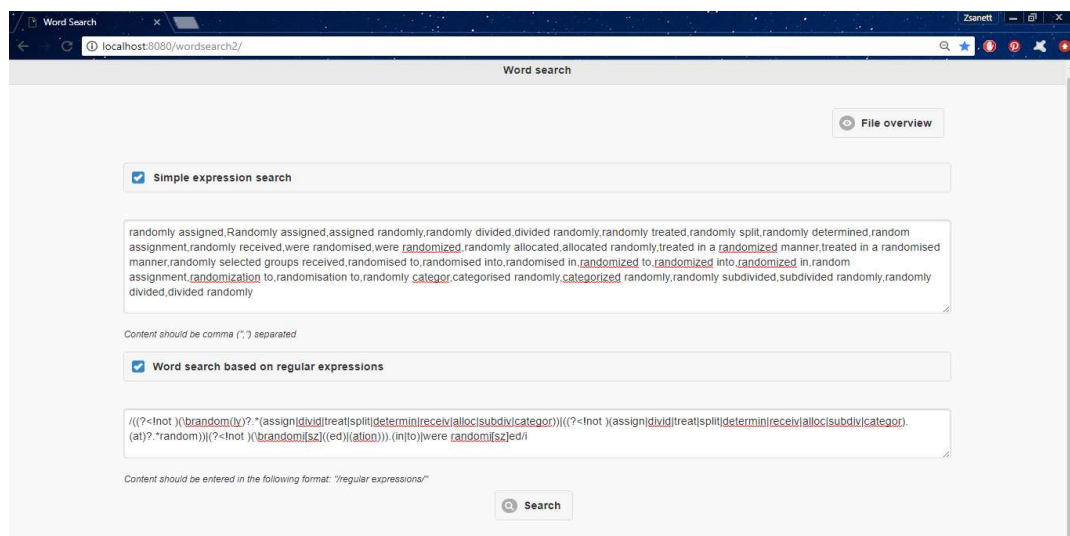


Figure 5.6 Program start page where user is able to perform two different types of searches

Based on the user's input, different workflows can be initiated - a simple expression search, a regular expression search, one where both are possible, and a file size check, the results of which can then all be used to update a database in the same way (Figure 5.7).



Figure 5.7 Four main workflows that use four different functions initially, and then share the same update database function

5.3.2.2.2 Results of Search page

Once the search has returned the results, these are displayed so that overall number of positive hits is shown as well as each publication where a hit has been identified is listed below with the corresponding piece of text identified within the file. At the top of the page, there is an option for the user to enter information relating to updating a database specified. This pilot project was designed to primarily update and tested on an MS Access database – a desktop relational database, but the application is designed to be able to work with other server-side relational databases as well (i.e. MSSQL, Oracle, MySQL). Here the user can specify the target table, the target column and the target value to be used for the Update Statement. The system by default will update all of the records that have come back with a positive result. The user is able to review and override this setting, by deselecting tick-boxes for records that they do not wish to include in the Update Statement (

Figure 5.8).

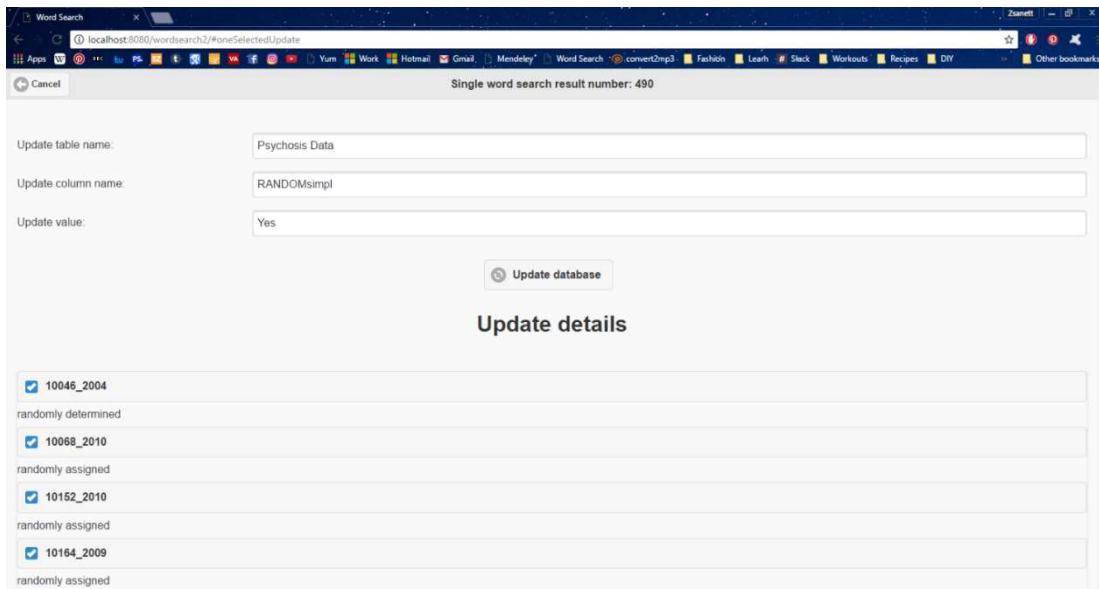


Figure 5.8 Appearance of the user interface when results of a single search are returned and displayed

As mentioned before, it is also possible to run both searches simultaneously, where the data to be displayed is selected by the user based on which calculation they are interested in, in a similar format to described above (i.e. results of both searches or just a single one) (Figure 5.9). This was mainly useful for a quick comparison of the two search approaches in finding the information of interest within the given set of files.

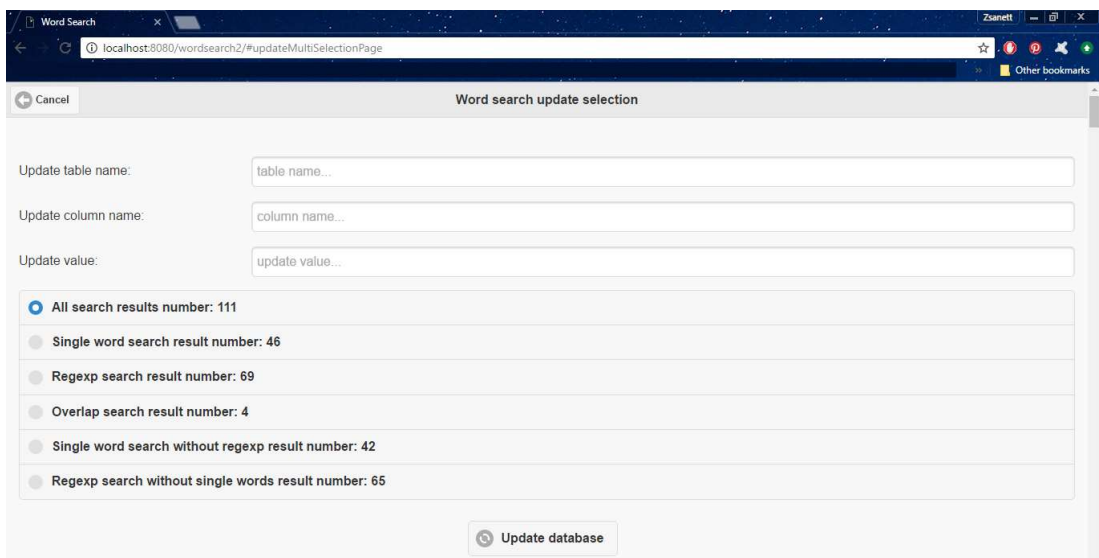


Figure 5.9 Appearance of the user interface when both searches are selected and executed simultaneously, and results are returned and displayed

5.3.2.2.3 Database update result page

The results of the executed Update Statement are displayed to show the overall success of the update process, including number of errors encountered, if any, and a log of all the records that had been successfully updated (Figure 5.10).

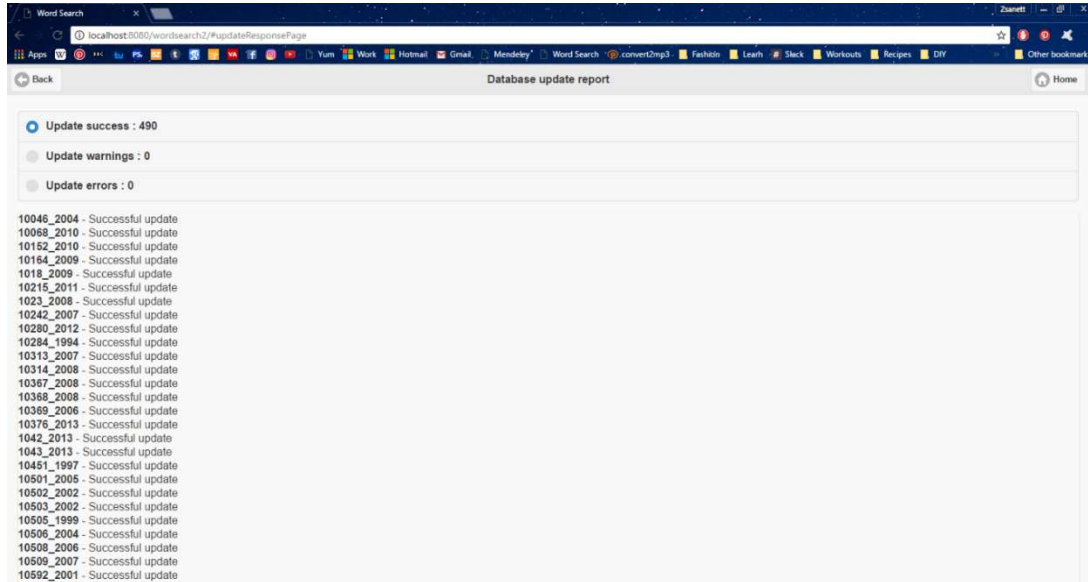


Figure 5.10 User interface as it appears after the user has updated the database

5.3.2.2.4 File overview page

Upon executing the 'File Overview' function on the Welcome page, the page changes to display the results of the file checker function by showing the total number of files that are likely to be erroneous (i.e. be incomplete in terms of text) and then also each file one by one contained within the folder of interest. This is calculated by setting a cut-off limit for the size of each file, where anything smaller than this limit is highlighted by the program and given a "warning" label. The program also highlights where no corresponding .txt file exists for a .pdf file. This way it is possible to exclude these records from final text mining analysis. It is also possible for the user to save this information to a database in a similar way to the way the results of the search were updated to a database (Figure 5.11).

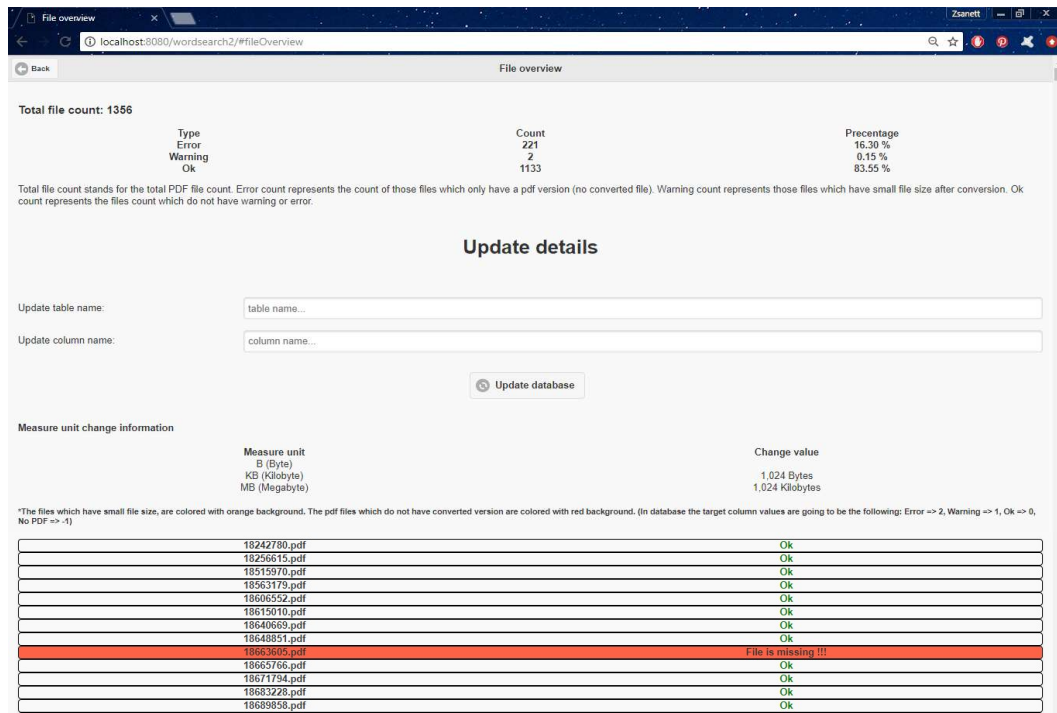


Figure 5.11 Results displayed on page in response to the program being prompted by the user to check files

5.3.2.3 Technology used to implement solution

For this pilot project, a web server and a back-end framework were set up. To create the web server the freeware Apache was used and for the back-end framework PHP (Hypertext Preprocessor) was used. The application was developed on two separate local environments at the same time – one on Windows and one on Linux operating systems – in order to design a portable and widely applicable solution that is compatible with both operating systems.

For front-end development, the following languages were used:

HTML (HyperText Markup Language) – used for the overall structure of user interface;

CSS (Cascading Style Sheets) – used for the styling and formatting of the HTML elements;

JS (JavaScript) – used for specifying interactions for dynamic data handling;

AngularJS – used to visualize dynamic data by simplifying the JavaScript code and HTML element visualizations;

JQuery - helper class for Javascript to aid access of HTML element values and attributions;

JQuery mobile – used to pre-define HTML elements in terms of styling to improve user experience.

A HTML document was used to specify the basic structural elements of the web page and thus define images, texts and inputs visible to the user. To support this, CSS was used, which is a HTML style sheet language essential in web design that allows the designer to specify the layout and the visual presentation of a web page by defining how HTML elements should be represented on screen. JavaScript, a dynamic scripting language was used for programming the behaviour of web pages. This was used to make the HTML document interactive. JavaScript is able to carry out simple functions and more specifically, we used JavaScript to generate various requests, for example, when a button is pressed, or when data are entered into fields, as well as for animation.

PHP was used to develop the backbone of the program and manipulate the operation of the program. This is an easy to use and widely used, server-sided scripting language that is especially useful for interactive and dynamic web development, and thus allows for the automated calculation of heavy code parts to run by a webpage hosting server. This programming language was chosen over some other ones because it is a good beginner language that is free, open source and easy to learn for beginner coders like me. It also has a strong support community behind it, which means that any bugs can be resolved fast, making the development process easier and quicker. Moreover, it is a clear and easy to understand language that can be edited in simple text editing software packages. Using this language, dynamic websites are achievable that are fast operating and can call on JavaScript objects and run on various operating systems, meaning the final product can easily be transferred and adapted to a new system.

As mentioned above, the front-end and the back-end communicated with each other through the posting of 'messages' written in JavaScript Object Notation (JSON) back and forth to each other. JSON is a data-interchange format, or in other words, syntax that is used to exchange and store data between the front-end and the back-end. When exchanging data between the two, data is required to be in a text format. JavaScript objects can be converted into JSON from the front-end and sent to the back-end, and converted back into JavaScript objects when JSON is sent back from the back-end to the front-end.

5.3.3 Development of Search words

5.3.3.1 Dictionary-based approach

Creation of search terms based on this approach were simply a list of phrases commonly found within our animal literature to correctly describe the risk of bias item in question. Possible variables within a phrase (i.e. species of animal mentioned or term used to describe experimenter) and thus different variations and deviations in word orders of a phrase had to be accounted for and added into the list of phrases as separate entries to create alternative options for wording of the same concept.

For randomisation, for example;

“animals were randomised to group”

would clearly not pick up on any of the following terms, despite being interpreted the same way by a reviewer:

“rats were randomised to group”

“mice were randomised to group”

Equally for blinding, the example below;

“investigators were blinded to treatment”

Would not positively identify the following:

“experimenters were blinded to treatment”

“observer was blinded to treatment”

Therefore, these options either had to be entered as separate items within my list, or as truncated expressions representing the shortest possible textual span that would identify as many correct phrases of interest as possible, while minimising identification of those that were not (i.e. “were blinded to treatment”). This latter approach could account for the high variability in syntax for the same basic concept. Of course, word order used between similar phrases limits this approach. In

addition, separate entries had to be considered for terms that might be affected by plurality, or letter case depending on whether phrase was at beginning or at the end of a sentence, and words, which might be in different forms of English (i.e. British English using “randomised” vs. American English using “randomized”).

5.3.3.2 Rule based approach

As this approach was rule based and specified a pattern, rather than exact phrases, it did not require entire phrases to be repeated for every single synonym or form of a word. Instead, ‘options’ could be added in for a certain phrase.

For example, the positive phrases mentioned above for randomisation;

“animals were randomised to group”

“rats were randomised to group”

“mice were randomised to group”

Could be picked up by the use of alternation and the rule:

(animals|rats|mice) were randomised to group

Where “|” denotes the option “or”

Similarly, where a letter might be different within a word (i.e. “s” vs. “z” in British or American English), this could be made into an option using regular expressions for character classes, where;

“animals were randomi[sz]ed to group”

*Where [..] are used denote where character class might be one or the other within
bracket*

Would match both:

“animals were randomised to group”

“animals were randomized to group”

This provides a powerful option to account for the large variety in syntax and synonyms. In addition, regular expressions can be set to ignore the case of words,

by using a “/” at the end of the list of expressions. This was another clear advantage over the previous approach, where cases of words would have to be accounted for using separate entries in our list of search phrases.

5.3.4 Performance of Solution and Evaluation

Overall success of the solution was measured by reassessing original objectives and seeing whether the system worked as intended.

5.3.4.1 Achievement of Objective 1: Creation of a set of term identifiers

The success of the prediction of reporting of measures taken to reduce the risk of bias through text mining was assessed by the calculation of specificity, sensitivity, precision and F-score. Scores could be interpreted so that when specificity is low it means that the computer falsely picks up on terms that would otherwise not be classed as an example of the reporting of the risk of bias item in question. In turn if sensitivity is low then the computer misses expressions that would otherwise be classed by a human as an example of the reporting of the risk of bias item in question. One would always indirectly influence the other and therefore the measurement of achievement would always involve a certain trade-off between the two values. Through discussions, we identified 80% as the minimum acceptable threshold value for specificity and sensitivity for the automation of risk of bias assessment, however, we also wanted a high sensitivity.

Through the dictionary-based approach, the computer called 112 publications as reporting randomisation to group, 188 reporting blinded assessment of outcome, and 7 reporting a sample size calculation. The rule-based approach identified 152 publications reporting randomisation, 187 reporting blinded assessment of outcome, and 37 reporting a sample size calculation. According to the human reviewers, 142 publications had reported randomisation, 192 blinding, and 8 a sample size calculation. The confusion matrices created when computer predictions were compared to human reviewers are displayed below (Table 5.1).

Table 5.1 Confusion matrices showing number of publication when comparing predictions made by the computer and classifications made by the human reviewers

Dictionary-based				Rule-based			
Randomisation		Human		Randomisation		Human	
		Yes	No			Yes	No
Computer	Yes	86	26	Computer	Yes	96	56
	No	56	966		No	46	936
Blinding				Blinding			
		Human				Human	
		Yes	No			Yes	No
Computer	Yes	170	18	Computer	Yes	170	17
	No	22	924		No	22	925
Sample Size Calculation				Sample Size Calculation			
		Human				Human	
		Yes	No			Yes	No
Computer	Yes	4	3	Computer	Yes	7	30
	No	4	1123		No	1	1096

5.3.4.1.1 Random allocation to group

Calling of the reporting of random allocation to group by the computer produces similar results when comparing the dictionary- vs. rule-based approach (Table 5.2 and Table 5.3). Sensitivity is low for both approaches, with dictionary terms resulting in 61% sensitivity and regular expressions performing slightly higher at 68%. Specificity on the other hand is very good for both, 97% for dictionary-based approach and 94% for rule-based approach. This means that for randomisation both approaches are good at ignoring papers that had not reported a risk of bias item within their experimental descriptions, but were less good at picking up on true examples of reporting. This is reiterated by the low value that we see for precision, and in turn the F-score. Precision of both approaches was similar at 77% for dictionary-based approach and 63% for rule-based approach. This means that the fraction of positives that were classified by the computer, over two-thirds were correct.

Table 5.2 Calculated performance measures for the search expressions developed using the dictionary-based approach

	Dictionary-based approach			
	Sensitivity/Recall	Specificity	Precision	F-score
Randomisation	61%	97%	77%	0.68
Blinded assessment of outcome	89%	98%	90%	0.89
Sample size calculation	50%	100%	57%	0.53

Table 5.3 Calculated performance measures for the search expressions developed using the rule-based approach

	Rule-based approach			
	Sensitivity/Recall	Specificity	Precision	F-score
Randomisation	68%	94%	63%	0.65
Blinded assessment of outcome	89%	98%	91%	0.90
Sample size calculation	88%	97%	19%	0.31

5.3.4.1.2 Blinded assessment of outcome

Automatic prediction of blinded assessment of outcome gives very good results across the board with both sensitivity and specificity returning percentages above 80%. For both approaches the tool performs at 89% sensitivity and 98% specificity, and gives a precision of 90% and 91% for dictionary- and rule-based approaches respectively. This also means that we get a very high F-score of 0.89 and 0.9 for dictionary- and rule-based approaches respectively.

5.3.4.1.3 Sample size calculation

The classification of publications according to whether a SSC had been reported performed very differently between the two approaches. Using the dictionary-based approach sensitivity of the tool was calculated at 50% and specificity at 100%. This implies that using this approach the tool picks up on a number of False Negatives and therefore misses publications where reporting evidence exists. Looking at the raw numbers, this is the case for both False Positives and False Negatives, although few in number. The reason why False Positives make more of an impact and result in a less sensitive tool, although not a less specific tool, is because of the rarity at which this measure is reported at within the literature. Therefore, the number of False Positives makes less of an impact when it comes to calculating Specificity as the number of True Negatives is large. On the other hand because the total number of Positives to be identified is low (i.e. 8), the percentage of False Negatives is a large proportion out of all the possible Positives. Precision is equally affected by these low numbers and in turn causes the F-score to be low too.

This also affects the rule-based approach, where even though the tool performs very well in terms of sensitivity and specificity, at 88% and 97%, respectively, precision is extremely low at 19%. This can again be explained by the low prevalence in the reporting of this methodological quality item.

5.3.4.2 Achievement of Objective 2: Creation of program to run and record results of term recognition

The success of the solution was measured against the initial user specification criteria (Table 5.4).

Table 5.4 Table showing level of success for each requirement that was specified at the start of the project

Pre-specified Software Requirements	Software's satisfaction of this criteria
User-friendly interface	✓ Simple to use and only contains key information
	✗ Limited guidance included on page about what display means - could be improved if used by additional users
Allow the investigation of the reporting of risk of bias item of choice	✓ Input is undefined and allows the entry of any string of search terms, thus not limiting risk of bias item to be searched
	✗ No option to select from pre-defined search terms, have to be entered manually
Search a number of different files at once	✓ Program directed onto folder specified within code - any files within this folder will be searched
	✗ Folder to be searched is currently burnt into the code of the program and cannot be redirected without editing the code
Facilitate the storage of these results in a database, if necessary	✓ Program updates MS Access folder, where table and column to be updated can be pre-specified
	✗ Can only currently update MySQL databases
Be possible to integrate into existing web-based platform	✓ As it is already designed as a webpage it can easily be moved to a web server and integrated into a web-based platform
	✗ Will need to adjust information about where program searches through files and what database it updates (i.e. database to be updated in future might also be web-based therefore requiring minor adjustments to code)

Overall, most of the points in my criteria were met to a standard where I, the primary user could use it for my research purposes. Limitations to the current solution would mainly arise if it were to be integrated into a web platform and used as a multi-user program.

This limitation, despite the program being a web-based solution that runs in a browser, is due to it being hosted locally and thus only accessible on the computer that it was designed on. For it to be accessible through the internet by multiple users, it needs to be hosted on a webserver or integrated into our existing web

platform (www.syrf.org.uk). For this to be then usable by anyone wishing to perform a risk of bias assessment of their dataset, some further developments would have to be made to the program. To fit in with the current platform the set of text files to be searched and their location would have to be redirected to the folder that contained them on our website. Furthermore, the database update process would need to be changed as the systematic review web platform uses a NoSQL database and will naturally store data slightly differently as the MS Access database that was used previously.

Finally, in the implemented program, I would like to create the added option of using previously developed regular expressions for users that may be less familiar with the reporting of risk of bias items. This will allow for both those new to systematic reviews and those more experienced in performing these sorts of analyses to use the current solution.

5.3.5 Critical analysis of results

5.3.5.1 Objective 1

My first objective was to develop a set of term identifiers that would pick up on a given set of phrases within a publication and automatically classify that publication according to reporting of risk of bias.

When classifying according to randomisation to group using the regular expressions approach, the computer identifies more publications positively than it does using the normal dictionary-based approach. This signifies that regular expressions allow for more freedom in syntax and variations to the same expression. Using the dictionary-based approach the computer identifies more false negatives, and using the regular based approach the computer identifies more false positives. Despite this difference, the two approaches still perform similarly because false positives have little impact, as the number of true negatives is so high. As a result, the proportion of false positives in relation to the total number of true negatives (i.e. used to calculate specificity) is very small and therefore specificity remains high. Therefore, while the dictionary-based approach here might be too strict in terms of only being able to identify a finite list of phrases, regular expressions can allow for too many options and pick up on examples that are very similar to those specified, but otherwise would not be included by a human investigator.

Classification of SSC worked well using the regular expression-based approach performing above our pre-specified success threshold for both specificity and sensitivity. It identified most possible examples of reporting of sample size calculation within the literature, but also identifying 30 more publications than the dictionary-based approach as reporting SSC. Therefore, it had a higher rate of false positives and thus very low precision. When using the dictionary-based approach, the sensitivity of the automated tool is low, and it misses a few instances of true reporting of SSC within the literature. On the other hand, specificity using this approach is very high. These dramatic differences in the observations in specificity and sensitivity measures of SSC are likely explained by the fact that the prevalence of reporting of SSC within a given field of pre-clinical literature is usually around 1% or less (Sena et al., 2014). This makes it very difficult to develop reliable and uniform regular expressions for this methodological quality item. It also explains why we see such a high percentage of specificity, as this takes into account the number of true negatives, which is high for these rarely reported items. On the other hand, sensitivity will be more affected by the false classification of a single publication as the proportion of publications reporting SSC is few. In fact when tested in a set of publications reporting experiments of animal models of lacunar stroke and middle cerebral artery occlusion (MCAO), we see a similar pattern where the regular expressions perform above the threshold for MCAO, but only perform with 50% sensitivity and 100% specificity for Lacunar stroke (Bahor et al., 2017).

When classifying publications according to blinded assessment of outcome, the tool performed above our target level of 80% for sensitivity and specificity in both dictionary and regular-expression approaches. It is likely that the classification of this risk of bias item by a computer had a better performance compared to randomisation and sample size calculation as blinded assessment of outcome, due to the fact that this is a more widely (Sena et al., 2014), and uniformly reported risk of bias item within the literature of many different pre-clinical fields.

5.3.5.2 Objective 2

The solution developed was able to speed up my work by both searching files, but also by being able to immediately update the desktop relational database Microsoft Access. The search of 1134 publications for the text relating to the reporting of a single risk of bias assessment took less than 2 minutes, and the updating of MS Access took a matter of seconds.

Having the program update a table in MS Access meant that I could immediately see and analyse the results of a search. This meant that using the tool I was able to get an overall prevalence of each risk of bias item in my validation dataset, as well as identify publications within this dataset where predictions made by the computer differed from human screener classifications.

The major limiting factor in the smooth running of the program was the conversion of pdf files to a readable text file format. Despite using different pdf converters, none were effective in converting 100% of pdf files to text files. Moreover, some converted files appeared to be converted but on closer inspection contained no usable information. This meant that these files had to be excluded from our set before analysis was run as otherwise they would have given false negatives, where the computer would be unable to identify an otherwise positive piece of text due to unavailability of the overall text. This problem was solved by adding in a "File checker" that would alert the user to files where no pdf existed to begin with, where pdf existed, but no corresponding text file could be found and those files that were unusually small and therefore unlikely to contain all the relevant information of the original pdf file.

5.3.6 Conclusions and Future Work

The aim of this project was to design a program that would speed up and potentially automate the annotation of the reporting of measures taken to reduce the risk of bias and other methodological quality criteria.

The tool created was successful from two aspects. First, it managed to reduce the time that it takes a human to score the reporting of measures taken to reduce the risk of bias. Classification by an experienced human reviewer is reduced from 5-10 minutes per publication (depending on the complexity of a publication) to about two minutes by an automated tool like this for the classification of 1134 publications per methodological criteria item. This will likely save on not only time of reviewers, but also resources such as additional reviewers. There is an opportunity for the computer to act as a second screener and reduce the need for additional personnel needed for systematic reviews. This is also the second strength of this approach. It is able to provide a method of checking the work of a human reviewer and thus reduce any impact of human error that might arise, as well as issues with subjectivity between different reviewers. While we maintain manual reviewing in the

field as the “gold standard”, this approach is not always superior and is just as likely to be at risk of things like human error. This has been shown in the field of law (Grossman and Cormack, 2011), but also by previous work from CAMARADES demonstrating impact of the implementation of machine learning for the screening of publications in a systematic review (Bannach-Brown et al., 2018).

While the solution is undoubtedly beneficial in many ways over a human reviewer, there are also some obvious limitations to replacing a human investigator with an automated system. Naturally, to decide the relevance of a particular document or a phrase within a piece of text human interpretation is required. There is often very little detail given about measures taken within an experiment to reduce the risk of bias, and it is not so straightforward what the correct interpretation of this evidence is. This makes it difficult to define specific rules and can lead to false positives and false negatives, where the algorithm given to the computer does not perfectly match that seen in the actual document. This is also the reason why the same tool might perform differently in some datasets when compared to others. At this stage, the tool is able to reliably automate the classification of publications for the reporting of blinded assessment of outcome. In some cases the classification of sample size calculation works well, but this along with the correct classification of reporting of randomisation to group needs more refinement before it can be confidently implemented in other reviews. It may be that there are differences between domains in the language that is used to describe risk of bias, and to make the tool applicable to a larger number of fields, this will need to be considered and refined in further iterations of the regular expressions used.

Despite these shortcomings, text mining seems to be a feasible approach for the automation of the classification of studies according to methodological criteria. It would also likely make a valuable contribution to our work. For example, it has long been established in law that human reviewers will disagree about the relevance of a piece of information in a large number of cases, independent of their level of expertise or detail to attention (Grossman and Cormack, 2011). Similarly in the sciences we recognise that in a systematic review single data extraction produces very different and much less accurate results than when two reviewers look at the same data and have a third independent investigator review discrepancies (Buscemi et al., 2006). This makes it clear that incorporating text mining in the process of systematic reviews can potentially not only save time and resources, but might also

in future be equally as good or even superior to human, “gold standard” data extraction and interpretation. Perhaps a solution for now would be to help aid a systematic reviewer’s work by providing a second reviewer for smaller reviews. For larger reviews, it is arguable to say that underestimating the prevalence of the reporting of a risk of bias item by 10% of the total prevalence is exceeded by the considerable amount of work and time that is eliminated and saved by automation of this process.

Here, I have only looked at three items off my methodological quality criteria list, but I believe with a bit more knowledge in the field, we can use text mining for other stages of the review process also. Mainly useful when it comes to categorization, but also potentially for some aspects of data extraction.

5.3.6.1 *Unsolved issues*

Some issues, which I was unable to account for during the development of this tool, included overcoming the issues with the conversion of pdf’s to text. While there are programs out there designed to perform this, they are all constrained by the nature of some of the pdf articles. For example, some pdf articles are ‘encrypted’, have restricted access for editing purposes, and therefore cannot be converted. Other articles might be scanned in copies, especially in the case of older articles and these cannot be “read” by a converter the same way as a modern pdf article. Moreover, a small proportion of articles would be converted using the tool, but would contain incomplete or disordered information from the source pdf (i.e. where text is displayed in multiple columns and sometimes these are not read in the right order during the conversion process). While the current tool attempts to exclude these files from the final analysis, these are all ongoing issues, which understandably limit the percentage of studies that can give a true estimate of the performance of the created tool.

At the stage of text processing, issues arose with syntax and the structure of text presentation. I accounted for case sensitivity of search terms and spelling variants (i.e. American vs. British spelling) in the regular expressions, but it was much more difficult to account for things such as hyphens that would break up an expression and potentially stop the tool from recognising the sequence of terms specified. Moreover, special characters might not be recognised by the converter and could get in the way of language processing. As the regular expressions presented here

were written in English, they could in turn only be used for English publications. In addition, the current solution did not specify the location where an expression had to be identified within a publication. This means that if a phrase appears within a title of a referenced article, this would be falsely classed by the computer as reporting a risk of bias item. Moreover, the tool did not take into account any information that was in addition to the main publication, such as supplementary materials published in separate files, that may have contained information relevant to experimental methodology and risk of bias. Such supplementary files are not commonly obtained and downloaded during the initial search when PDF's are collected for publications and therefore cannot be used to inform the final classification results. A quick search in the psychosis dataset shows that supplementary information is referred to in about 25% of the publications, included within the review and about 16% of publications in the test dataset refer to any form of supplementary information. Furthermore, the reporting of measures taken to reduce the risk of bias and other methodological criteria are recommended by reporting guidelines to be covered in the main text.

Finally, concept drift might affect the effectiveness of the search terms developed at current, which means that as language evolves, terminology used to describe some methodological criteria might change over time and thus will require further human input to update search terms.

5.3.6.2 Limitations to text mining as a tool for methodological criteria assessment

Obviously, the effectiveness of the system is based on the knowledge that is fed to the computer in order to aid its decision-making. This, now, is purely based on the occurrence of pre-specified phrases and therefore will be limited if the computer comes across any phrases which deviate from this list of phrases. This can be improved by increasing the breadth of "knowledge" that is supplied to the computer in the form of the search phrases used by testing and validating in increasing amount of datasets. This increases our awareness of examples used to describe the same concept in the literature that search terms are based on. Therefore, the more data analysed, the more accurate the model will be.

In addition, the number of publications which text mining can be applied to is limited by the above-mentioned limitations of issues with file conversion and language

restrictions. These drawbacks of text mining all limit the application of this process as well as the complete replacement of a human investigator.

5.3.6.3 Further development plans

Currently, the regular expressions I created through this project for the rule-based approach have been implemented into our systematic review platform SyRF (Figure 5.12). It does not implement the program created, because SyRF currently has a different infrastructure and does not require the database update function. Perhaps other features of the program can be integrated into the platform, but at its current state, this is still under development.

Title	Year	Reference Type	Journal	Author	Inclusion Status	Search	Blinding	Randomisation	Sample Size
Electroacupuncture-like stimulation at the Bailhui (GV20) and Dazhui (GV14) acupoints protects rats against subacute-phase cerebral ischemia-reperfusion injuries by reducing S100B-mediated neurotoxicity	2014	Journal Article	PloS one	C. Y. Cheng	Included	1st	No	No	No
Protective Effects of Dihydrocaffeic Acid, a Coffee Component Metabolite, on a Focal Cerebral Ischemia Rat Model	2015	Journal Article	Molecules (Basel, Switzerland)	K. Lee	Included	1st	Yes	No	No
Ischemia induces different levels of hypoxia inducible factor-1alpha protein expression in interneurons and pyramidal neurons	2014	Journal Article	Acta neuropathologica	P. Ramamoorthy	Excluded	1st	No	No	No
Myeloperoxidase propagates damage and is a potential therapeutic target for	2015	Journal Article	Journal of cerebral blood flow and metabolism : official journal of the International Society of	R. Forghani	Included	1st	Yes	No	Yes

Figure 5.12 Screenshot of how automated assessment of measures taken to reduce the risk of bias is currently being implemented in the online platform SyRF (www.syrf.org.uk) – screenshot taken 22/08/2017

Overall, the tool as it is, has been beneficial for my personal use and will continue to allow me to develop and test other search strings for other methodological criteria items as well as refine those that have already been designed. So for my own use, it would be interesting to take the program further and not just display phrases found, but have an option where the user can choose to look at the file of their choice and

be shown where the phrase has been identified. This can further be improved by allowing the user to manually edit the outcome for that publication and thus update the overall total. This might not be such a feasible option for large reviews but would create the optional opportunity for the user to check what the computer has done. Moreover, incorrect predictions by the computer could be accumulated to help further refinement of the expressions used for term recognition. Finally, an important limitation of the current tool is that it combines studies that explicitly said they did not take a measure to reduce risk of bias (i.e. did not randomise), with studies that might not have otherwise given enough information to determine whether they had or not. These were collectively labelled as no evidence of reporting of measures taken to reduce the risk of bias. This could potentially be further refined in future iterations of the program, to be classed as separate judgements. This would be in line with tools such as SYRCLE's RoB tool, which makes this differentiation. In addition to "Yes" and "No" indications for low and high risk of bias, respectively, there is also an option to assign a judgement of "unclear" risk of bias to each item, where there is not enough information to determine risk of bias (Hooijmans et al., 2014).

6 Methodological Quality in the Literature

6.1 Introduction

Research is informative and of high-quality, when experiments are well-designed in advance, they are rigorously carried out during, and results obtained are analysed correctly afterwards (Samuel et al., 2016). In line with methodological quality assessment is the consideration of the extent to which a study is at risk of bias within an experiment by looking at measures taken within studies to reduce these (Krauth et al., 2013). Risks of bias can lead to systematic errors (i.e. deviations from the “truth”) in the results that we see in studies. They can lead to overestimation or underestimation of a seen effect (Higgins and Green, 2008). The well-known study by Rosenthal and Fode in the 1960s demonstrated this concept very well, showing that experimenter bias in his students could in fact have a large impact on the differences in results, obtained in a study looking at the performance of the same groups of rats in a learning and memory paradigm (Rosenthal and Fode, 1963).

While the importance of methodological quality (Chalmers et al., 1981) and quality of reporting of important methodological considerations, such as randomisation and blinded assessment of outcome (Begg et al., 1996), have long been established for clinical research, the animal research field seems to be lagging behind (Landis et al., 2012). In recent years it has become repeatedly clear that results from animal studies are affected by similar methodological criteria (Bebarta et al., 2003) and that the reporting of these measures within the literature of different fields of animal research is generally poor (Macleod et al., 2015).

Internal validity in animal studies may be affected by four basic types of bias, which have the potential to introduce systematic differences between experimental groups within a study (van der Worp et al., 2010). These include 1) selection bias, when animals are allocated in a biased manner to a treatment group – overcome by randomisation and concealing allocation of animals to groups; 2) performance bias, when care and handling of animals differs between groups outside of intervention under investigation – overcome by blinded assessment of outcome; 3) detection bias, which is the systematic distortion of results as a result of investigator having knowledge of treatment assignment of animals – overcome by blinded assessment of outcome; and finally 4) attrition bias, when systematic differences arise between groups through the reporting of incomplete data as a result of omission or exclusion

of animals from a study – overcome by blinded assessment of outcome and intention-to-treat analysis (van der Worp et al., 2010). Randomisation involves randomly allocating animals to an intervention group so that they have an equal chance of being allocated to either treatment or control group, while allocation concealment means concealing this group assignment sequence from those that will perform this allocation process (Higgins and Green, 2008; van der Worp et al., 2010).

In addition to bias, imprecision, which describes the likelihood of random error is also important and commonly tested for in systematic reviews (Higgins and Altman, 2008). While these do not put a study at risk of bias per se, smaller studies are less precise and therefore could potentially falsely miss important biological effects that are otherwise present (Krauth et al., 2013; Landis et al., 2012). Moreover, selective reporting bias (discussed in greater detail in the next chapter, Chapter 7), is addressed by the good scientific practice of designing a protocol for a study before it is carried out, and subsequently making this available so that all analyses originally intended are performed and there is no cherry picking of data.

Other factors to consider when evaluating the quality of a published study are certain reporting criteria, which can affect the outcomes within a study. These include the disclosure of any potential conflicts of interest and compliance with animal welfare regulations. In clinical studies, for example, it has been shown that if studies and investigators have financial ties, then research outcomes are usually supportive of funders (Lundh et al., 2017). Finally, animal welfare is not only recognised to be important in obtaining reliable results in the laboratory (Poole, 1997). It has been shown that keeping high standards of animal welfare in laboratory experiments is important for the validity of animals as models, for the disorders in question and in order to make sure that studies are reproducible (Prescott and Lidster, 2017).

While both lumped together here within this project under “internal validity”, it is important to differentiate between ‘quality’ and ‘risk of bias’. Performing a study to a high quality, does not necessarily mean that the study is free of risk of bias (Higgins and Altman, 2008).

Ultimately, it has been suggested that poor methodological quality of animal experiments may impede the translation of results from this domain of research to

humans in a clinical setting (Hooijmans and Ritskes-Hoitinga, 2013). Therefore, it is of interest whether this could also be to blame for the difficulties in translation of animal research to new drug developments in the clinic within the psychotic disorders research field. For this reason, I was interested in looking at the prevalence of the reporting of certain study design criteria, including risk of bias, methodological and reporting criteria, in the pre-clinical literature of psychotic disorders. In this chapter I report on this overall prevalence in the literature, how it has changed over time and explore whether certain study characteristics might affect the reporting of these measures in published studies.

6.2 Methods

During the categorisation phase of the project (phase II) I looked at scoring publications against a pre-specified list of 8 study design criteria, which included items of risk of bias, methodological and reporting criteria. Some publications were manually categorised against these pre-specified criteria, and the rest of the publications were categorised for three of the items in the list using methods described in detail in the previous chapter, Chapter 5. These items were random allocation to group, blinded assessment of outcome and sample size calculation.

To test for differences in reporting based on study design, I tested for equality of proportions of subgroups using the proportion test, calculated using the `prop.test` function in R, which uses the Chi-squared test for independence. As the chi-squared approximation is affected by small sample size, I performed Fisher's exact test using the function `fisher.test` in R, as a sensitivity analysis. Due to the testing of associations between 8 risk of bias items and different methodological factors the critical value of p was adjusted to 0.006, using Holm-Bonferroni correction.

When the chi-squared test proved to be significant for tests involving more than two samples, data were further investigated using all possible pairwise comparisons with Bonferroni corrections of p values (MacDonald and Gardner, 2000). This meant that for the 4-sample tests a corrected p value of 0.008 (i.e. $0.05/6$ for 6 possible pairwise comparisons) was used.

6.3 Results

6.3.1 Overall reporting of risk of bias and other quality measures by studies in the literature

During Phase II of the project, 3847 publications were included and categorised. I was able to categorise 2462 (64%) publications of this set manually for reporting of measures on my list of quality criteria, while 1387 publications were categorised using the text mining tool developed by me. The studies, which were manually categorised, were entirely random and there was no obvious difference in terms of study design or publication date between studies manually categorised and those categorised by the computer.

Overall, of 2462 publications categorised manually, 573 (23%) reported randomising animals to group, 49 (2%) reported concealing allocation of these group assignments from investigators and 613 (24.9%) reported performing assessment of outcomes blinded (Table 6.1). Only 7 (0.3%) reported to have carried out a sample size calculation, and equally as few 5 (0.2%) publications reported the availability of a study protocol for their studies. In terms of reporting criteria, 1943 (79%) reported that their experiments were approved by and complied with animal welfare regulations, 726 (30%) included a statement of potential conflict of interest, including whether there was any to disclose or not. Overall, median score for number of items reported from the list of publications was 2/8.

Table 6.1 Overall prevalence of the reporting of risk of bias items and other methodological quality criteria in the field

	Categorised Manually	Categorised by Computer
Total No. of publications categorised	2462	801
Number of publications reporting ... (Overall %)		
Random allocation to group	573 (23.3%)	158(19.7%)
Allocation concealment	49 (2%)	-
Blinded assessment of outcome	613 (24.9%)	162 (20.2%)
Sample size calculation	7 (0.3%)	27 (3.3%)
Compliance with Animal Welfare Regulations	1943 (78.9%)	-
Statement of potential conflict of interest	726 (29.5%)	-
Exclusion of animals pre-specified or explained	416 (16.9%)	-
Availability of a study protocol	5 (0.2%)	-
Median quality (/8)(interquartile range)	2 (1-2)	-

The text mining tool developed for randomisation, blinded assessment of outcome and sample size calculation was run on the rest of the publications that could not be manually categorised. Of 1387 publications, 801 publications could be converted for text mining utilising the text converter. Overall prevalence of reporting was equally as low here among publications. 158 (20%) of publications reported randomising animals to control or treatment groups, 162 (20%) of publications reported assessing outcomes in a blinded manner and 27 (3%) of publications reported the performance of a sample size calculation.

Based on publications ascertained for these items manually, the tool performed at a sensitivity of 80% for randomisation, 83% for blinding and 100% for sample size calculation. Levels of specificity were calculated to be 95% for randomisation, 91% for blinding and 98% for sample size calculation in this dataset. This means that the tool seems to slightly underestimate the true prevalence of the reporting of random allocation to group, and blinded assessment of outcome in the literature, and overestimates the prevalence of the reporting of sample size calculations in the literature. This is also shown by the numbers in the manually categorised dataset and the dataset that has been categorised by the computer.

Nevertheless, by aggregating the results from the two approaches I found that random allocation of animals to group is reported in about 22.4% of publications describing animal models of psychotic disorders, while blinded assessment of outcome is reported in about 23.8% of publications and sample size calculations is reported in about 1% of publications.

6.3.2 Change in prevalence of reporting of risk of bias and other quality measures as a result of different factors

As I could only correctly ascertain all items on my methodological quality list for studies that had been manually categorised against the checklist, all further analyses were done using the manually categorised set of included studies ($n = 2462$).

6.3.2.1 Change in prevalence of reporting over time

In recent years there has been increasing focus on the importance of better reporting of animal experiments including increasing transparency around measures taken to reduce risks of bias within animal studies and of the reporting of other methodological criteria which might help increase predictive validity of these preclinical studies (Landis et al., 2012).

In this context, I was interested in seeing whether the reporting of these risk of bias items and other methodological criteria had changed in the literature over time. Taking publications included and categorised in Phase II of the review I looked at the total prevalence of reporting of these measures over the years of publications of these studies.

It appears that the reporting of these measures has increased over time especially since the early 1990s (Figure 6.1). The sharpest increase is seen in the reporting of compliance with animal welfare regulations and the inclusion of a statement of potential conflict of interest, whereas there doesn't appear to be a significant increase in the reporting of a sample size calculation and availability of a study protocol among published studies. The reporting of allocation concealment appears to have very slightly increased in recent years. Across publications in the literature, the reporting of random allocation of animals to group, blinded assessment of outcome, and accounting for all animals through the reporting of exclusion of animals are all items that seem to have gradually increased over time.

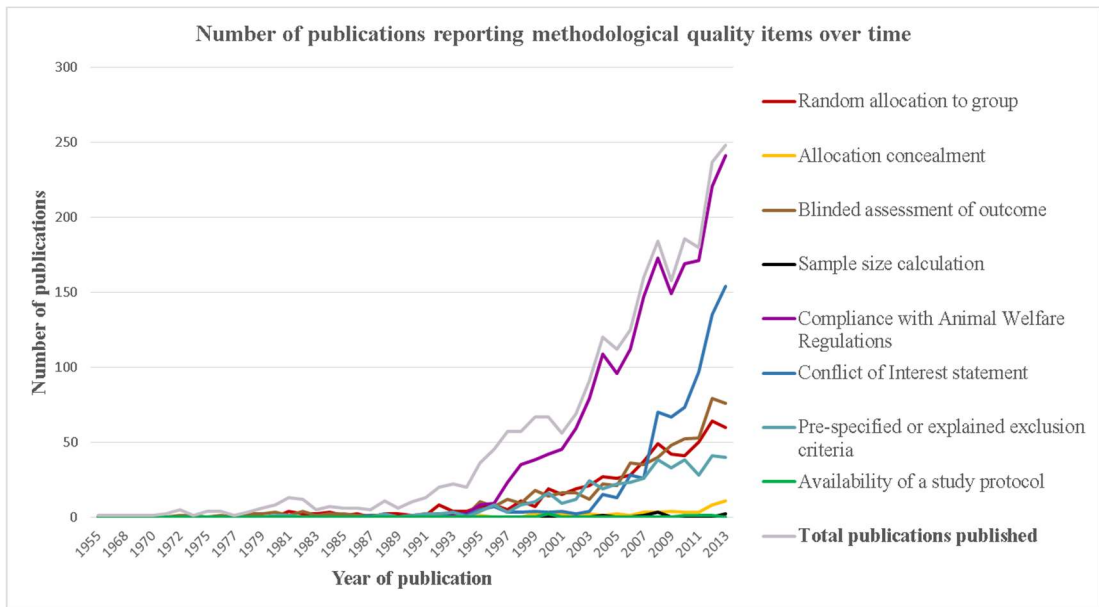


Figure 6.1 Prevalence of reporting of risk of bias items and other methodological quality criteria in the field over time

Nevertheless, it must also be taken into account that the number of publications published every year has also increased, which could make these observations slightly misleading. Figure 6.2. shows the true prevalence of reporting of these study design measures each year.

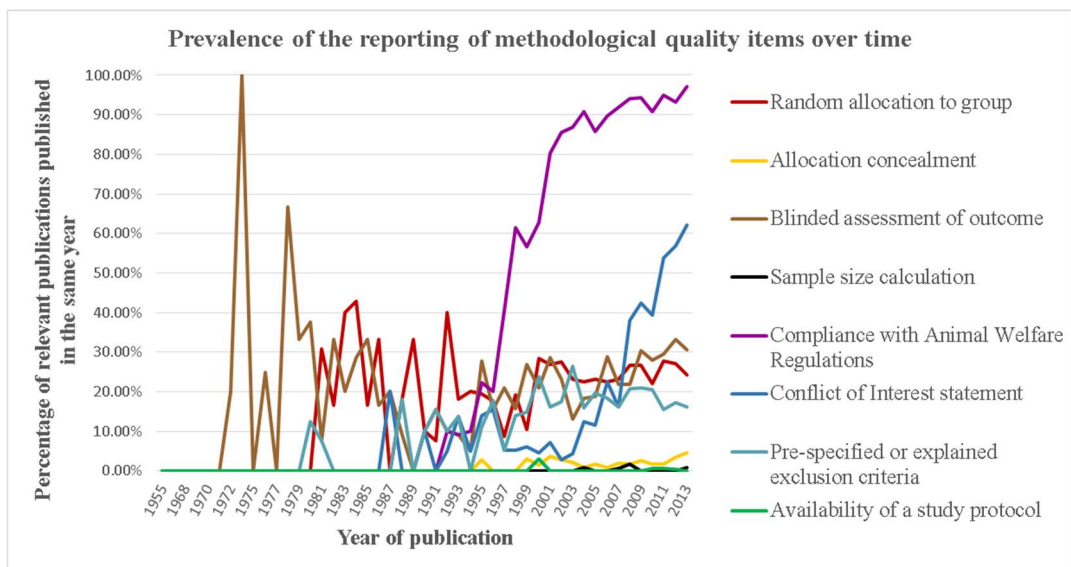


Figure 6.2 Prevalence of the reporting of risk of bias items and other methodological quality criteria as a percentage of total publications reported in the same year

This apparent increase in the overall prevalence of the reporting of risk of bias items such as random allocation of animals to group and blinded assessment of outcome has actually been very gradual if at all different compared to earlier years. It is also arguable based on this graph that the prevalence of reporting of these two measures in the literature has varied the least over the years. We still see a stark increase in the prevalence of the reporting of compliance with animal welfare regulations and the inclusion of a statement of any conflicts of interest. It is also apparent that the reporting of allocation concealment in the literature has not changed substantially since the late 1990's. The reporting of animal exclusions looks to have decreased over the last decade and the availability of a study protocol and sample size calculations have continued to be poorly reported.

6.3.2.2 *Change in prevalence of reporting in types of studies*

In clinical trials for novel interventions, risks of bias and measures taken to reduce these are recognised to be an important factor in minimising the number of wrong conclusions that are reached about the efficacy of an intervention (Gluud, 2006). In this context, I was interested to see whether the type of study carried out affected the reporting of study design items and whether drug studies were more likely to report these measures as a direct result of clinical trial practice over model characterising studies.

Manual categorization identified 1124 publications measuring the effect of an intervention in animal models of psychotic disorders that were administered in order to prevent or reverse effects of the model. The remaining 1338 publications were classed as model characterising studies comparing healthy or sham animals to animals that had an intervention to model an aspect of a psychotic disorder.

Overall, dividing studies manually categorised for study quality showed very little difference on the prevalence of reporting of measures within the list of methodological criteria items (Figure 6.3). Only significant difference was seen in the reporting of a statement of potential conflict of interest, whereby the proportion of treatment exploring studies that reported a statement of potential conflict of interest was greater than the proportion of model characterising studies reporting it ($\chi^2 = 10.23, p = 0.0014$).

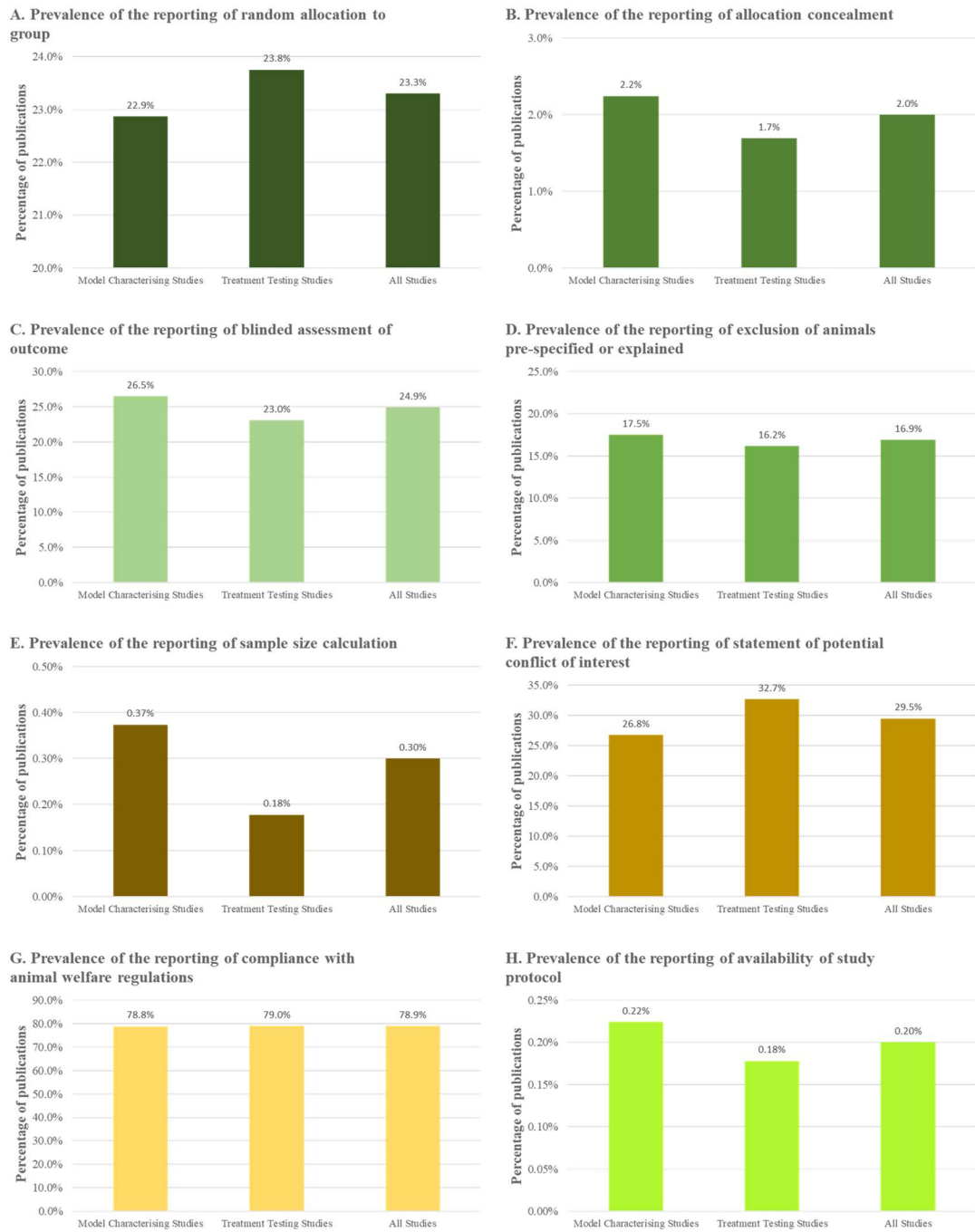


Figure 6.3 Prevalence of the reporting of each item on the list by type of study

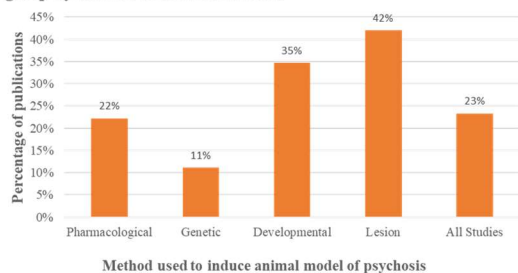
6.3.2.3 Change in prevalence of reporting across publications using different models

As I have shown in Chapter 3, animal models of psychotic disorders can be largely grouped into 4 different methods of model induction. Manual categorization of included studies identified 1725 pharmacological (including substance studies), 352 genetic, 692 developmental and 88 lesion studies. Naturally some publications can use more than one method to induce models, with many using these in combination. This factor could not be taken into account here but might mean numbers seen here for one method of induction were confounded by another method reported together with it in some publications.

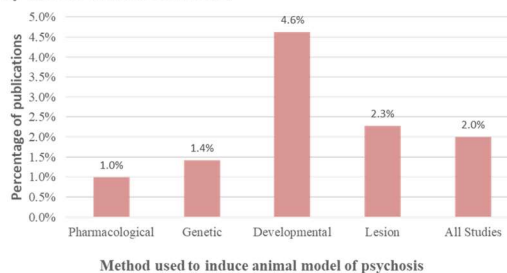
When looking at method of model induction individually, data shows that the proportion of studies reporting random allocation to group is significantly different depending on the method of model induction reported within a study ($\chi^2_3 = 92.71$, $p = 2.2 \times 10^{-16}$). The proportion of genetic studies reporting this item was significantly less than the proportion of all other studies reporting this risk of bias item (39/352, 11% of publications reporting genetic models, Figure 6.4, $p < 2 \times 10^{-5}$ for all pairwise comparisons between methods used). In contrast, this risk of bias item was significantly more widely reported in studies reporting lesion models (37/88, 42% of publications reporting lesion models, proportion of lesion studies reporting randomisation significantly higher than proportion of genetic, $p = 1.1 \times 10^{-10}$, and pharmacological studies, $p = 0.0002$), and developmental models (240/692, 35% of publications reporting developmental models, proportion of developmental studies reporting randomisation significantly higher than proportion of genetic, $p = 4.1 \times 10^{-15}$, and pharmacological studies, $p = 1.9 \times 10^{-9}$). Reporting of allocation concealment was in general poor across all studies, but the proportion of studies reporting it was significantly different depending on the type of model used, despite small sample sizes ($\chi^2_3 = 34.66$, $p = 1.4 \times 10^{-7}$, significant results confirmed by sensitivity analysis using Fisher's exact test). It was most prevalent among studies reporting developmental models (32/692, 5% of publications reporting developmental models, proportion of developmental studies reporting allocation concealment significantly higher than proportion of pharmacological studies, $p = 1.5 \times 10^{-7}$, no other significant differences between pairwise comparisons). Blinded assessment of outcome was reported significantly differently across different studies ($\chi^2_3 = 27.46$, $p = 4.7 \times 10^{-6}$). It was most widely reported by publications reporting genetic and developmental

models (127/352, 36% of publications and 184/692, 27% of publications, respectively). Interestingly pharmacological studies reported this item even less frequently than all studies combined did (399/1725, 23% of publications reporting pharmacological models compared to 25% of all studies, proportion of pharmacological studies reporting blinding significantly lower than proportion of genetic studies, $p = 3 \times 10^{-6}$, no other significant differences between pairwise comparisons). The reporting of any exclusion of animals was highest in studies reporting lesion models (39/88, 44% of publications reporting lesion models, proportion of lesion studies reporting exclusion of animals significantly higher than proportions of all other types of studies, all pairwise comparisons $p < 9.2 \times 10^{-6}$ for lesion studies). There was little difference in the prevalence of reporting between other studies of the same measure and the overall prevalence (17%). Sample size calculation was equally poorly reported among all studies irrespective of the method of model induction reported (no significant differences). Studies reporting genetic studies included a conflict of interest statement (168/352, 48% of publications reporting genetic models) significantly more commonly than studies using other methods of model induction ($\chi^2 = 58.46$, $p = 1.25 \times 10^{-12}$ overall, and $p < 9.6 \times 10^{-5}$ for all pairwise comparisons for genetic studies). Genetic studies also reported compliance with animal welfare regulations (312/352, 89% of publications reporting genetic models) significantly more commonly than pharmacological studies ($p = 6.4 \times 10^{-7}$). This item was also well reported in studies describing developmental models (604/692, 87% of publications reporting developmental models, proportion of developmental studies reporting compliance with animal welfare regulations significantly higher than proportion of pharmacological studies, $p = 1.4 \times 10^{-9}$, no other significant pairwise comparisons). And finally, the availability of a protocol for the study included in the review was rarely mentioned within studies, which was not affected by the method used to induce the model described.

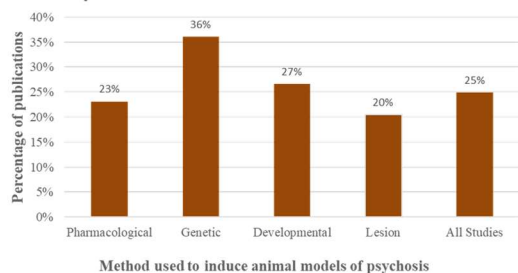
A. Prevalence of the reporting of random allocation to group by method of model induction



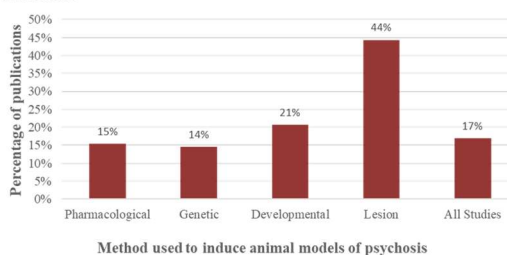
B. Prevalence of the reporting of allocation concealment by method of model induction



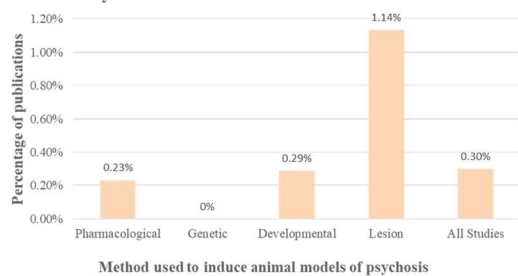
C. Prevalence of the reporting of blinded assessment of outcome by method of model induction



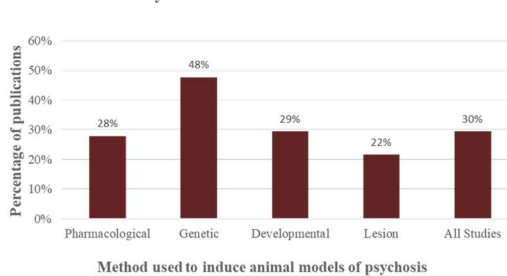
D. Prevalence of the reporting of pre-specified or explained exclusion of animals by method of model induction



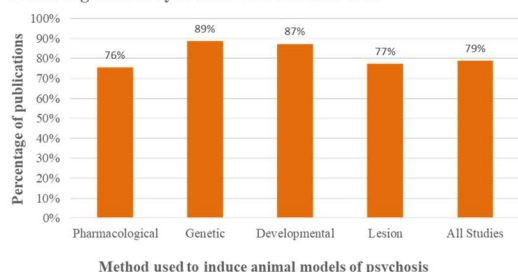
E. Prevalence of the reporting of a sample size calculation by method of model induction



F. Prevalence of including a statement of potential conflict of interest by method of model induction



G. Prevalence of reporting of compliance with animal welfare regulations by method of model induction



H. Prevalence of the reporting of availability of a study protocol by method of model induction

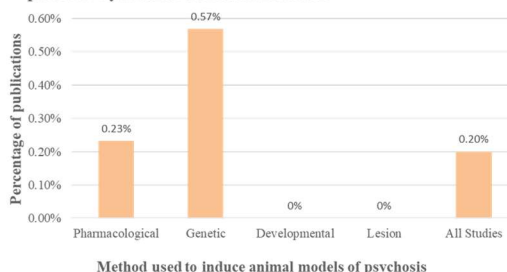


Figure 6.4 Prevalence of the reporting of each item on the list by method used to induce model

6.3.2.4 Change in prevalence of reporting in publications reporting different outcome measure types

During the categorization phase of the project, publications were classified according to the type of outcome measure that was reported. I differentiated between behavioural, anatomical, neurochemical and electrophysiological outcome measures. I wanted to see whether any of these outcome measures would influence the prevalence of publications that reported each item of interest on my list of methodological criteria.

In total, 1837 publications were recorded as having measured behavioural outcomes, 288 as having measured anatomical outcomes, 1120 as measuring neurochemical outcomes and finally 273 publications as measuring electrophysiological outcomes. It is important to keep in mind that some studies measure more than one type of outcome measure and therefore, results for one type of outcome measure might be confounded by the effect of another outcome measure in publications that looked at more than one outcome. Overall, I found that prevalence of reporting of study quality measures was equally low across all studies, irrespective of type of outcome measured (Figure 6.5). Nevertheless, the proportion of studies reporting randomisation was significantly different based on the type of outcome measured ($\chi^2 = 35.22$, $p = 1.1 \times 10^{-7}$), whereby the reporting of this risk of bias item was least prevalent among studies measuring electrophysiological outcomes (29/273, 11% of publications measuring electrophysiological outcomes, proportion of electrophysiological studies reporting randomisation significantly lower than proportion of behavioural studies, $p = 6.8 \times 10^{-7}$, and neurochemical studies, $p = 0.0002$). Allocation concealment was poorly reported across all studies, but was most widely reported in studies measuring electrophysiological and anatomical outcomes (both 4% of publications measuring each outcome measure compared to 2% of all publications, no overall significant differences between outcome types). Blinded assessment of outcome was most widely reported in publications reporting measuring anatomical outcomes (140/288, 49% of publications measuring anatomical outcomes, proportion of anatomical studies reporting blinding significantly higher than proportion of all other outcome type studies, all pairwise comparisons $p < 4 \times 10^{-12}$ for anatomical studies), and least commonly reported among studies measuring electrophysiological outcomes (53/273, 19% of publications measuring electrophysiological outcomes, only significantly different in comparison

to proportion of anatomical studies). The reporting of exclusion of animals differed significantly between type of outcome measure used ($\chi^2_3 = 23.86$, $p = 2.7 \times 10^{-5}$), whereby the proportion of neurochemical studies reporting this item was significantly lower than behavioural studies ($p = 0.0001$) and electrophysiological studies ($p = 0.0007$). Sample size calculation was rarely found to be reported in any of the studies included in the dataset with no significant difference between studies reporting different outcome types. There was little variation in terms of publications including a conflict of interest statement or the reporting of compliance of experimental methods with animal welfare regulations, across publications measuring different outcome measure types (no significant differences between studies reporting different outcome types). Equally, the reporting of the availability of a study protocol was poor across all studies, with little effect of the type of outcome that was reported to be measured within studies (no significant differences between studies reporting different outcome types).

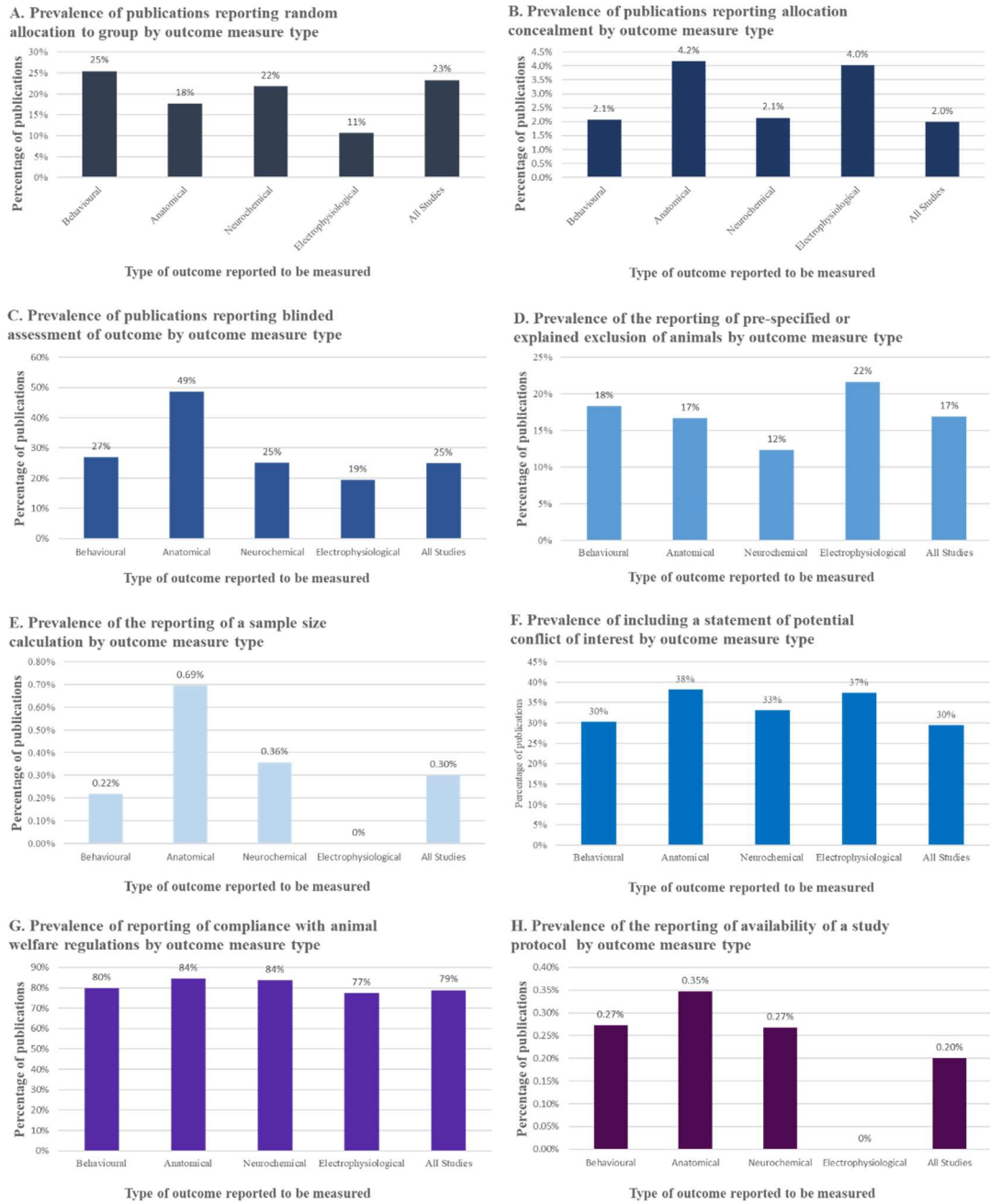


Figure 6.5. Prevalence of the reporting of each item on the list by outcome measured

6.3.3 Does low quality of reporting affect outcome reported within publications?

Using data extracted as part of a subset of studies within this review, which describes animal experiments of developmentally induced models of schizophrenia and is described in more detail in Chapter 4, I looked at how these trends in reporting affect the outcomes that are reported. None of the differences between studies reporting or not reporting measures taken to reduce the risk of bias were statistically significantly for model characterising studies (Figure 6.6). This included looking at the reporting of random allocation to group, allocation concealment, blinded assessment of outcome and exclusion of animals. I was unable to look at the difference between studies reporting or not reporting sample size calculations, despite this being of high interest, as data was insufficient for analysis for studies that did report this measure.

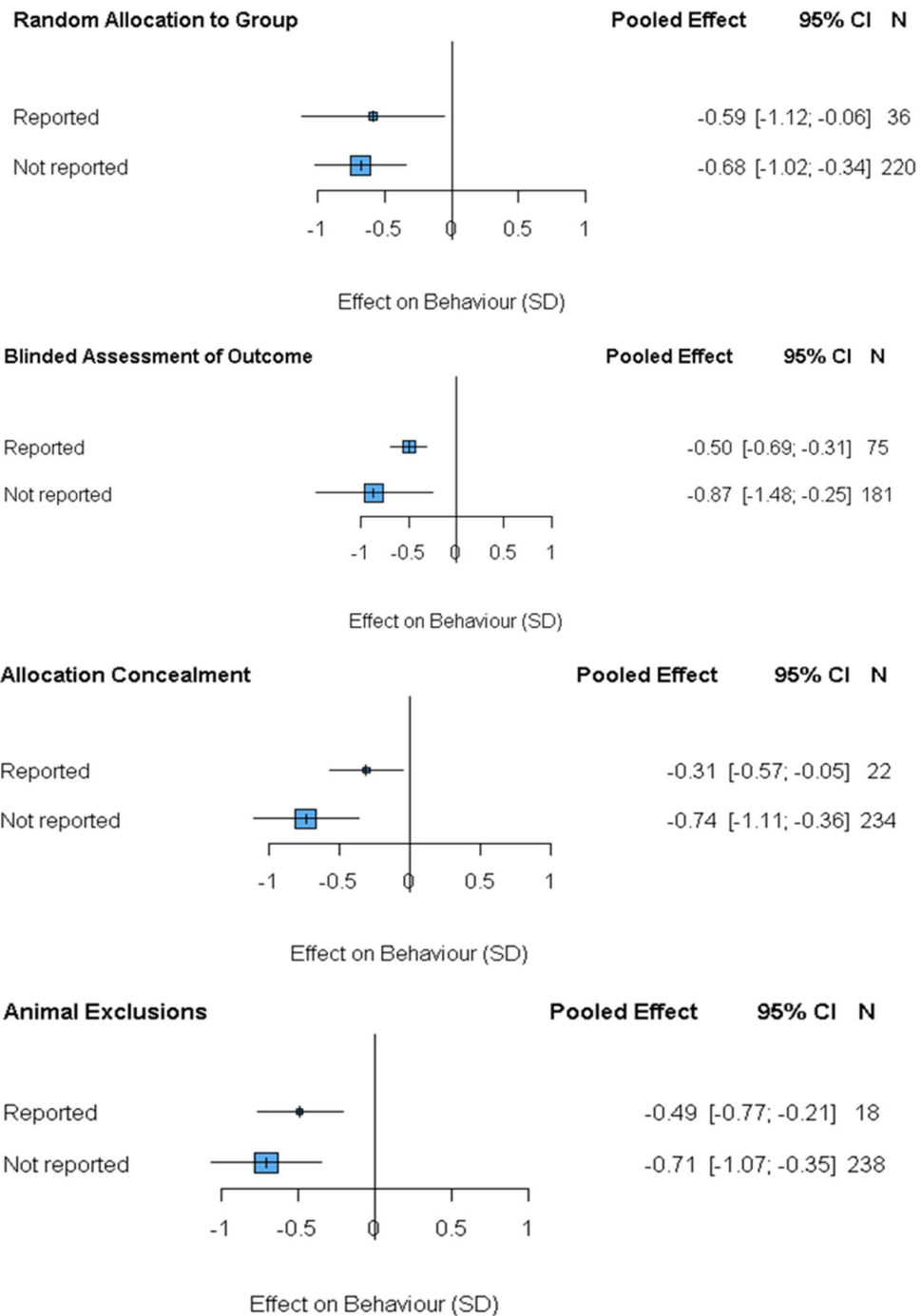


Figure 6.6 Model characterising studies reviewed here that do not report certain risk of bias items do not show statistically significantly different effects when compared to studies that do report these risk of bias items.

When looking at treatment testing studies, the reporting of random allocation to group significantly affected the efficacy of treatments reported within a study, whereby those that reported this risk of bias item also reported lower efficacy of treatments on model animals, compared to studies that did not report this risk of

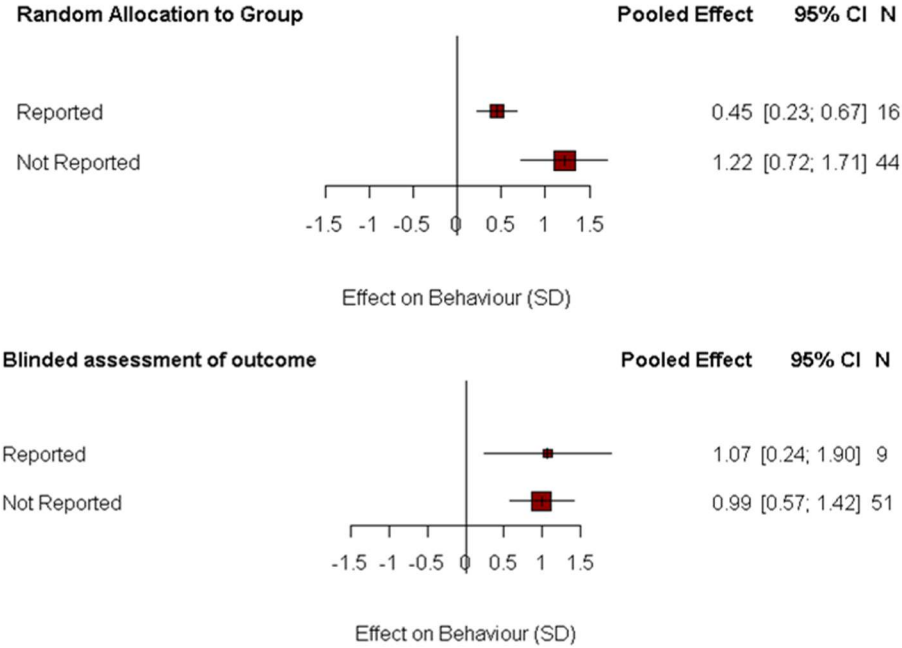


Figure 6.7 Studies that do not report random allocation to group overestimate the efficacy of treatments compared to studies that do report it. In comparison, the effect of studies reporting or not reporting blinded assessment of outcome were not found to be statistically significantly different.

bias item (Figure 6.7). There was no significant difference seen in treatment effects between studies that reported blinded assessment of outcome and those that did not. Not enough treatment testing studies reported a sample size calculation, allocation concealment or animal exclusions in order to assess the impact of reporting these items in treatment testing studies.

6.4 Discussion

Overall data shows that the prevalence of reporting of measures taken to reduce the risk of bias, good methodological and other important reporting criteria are poorly reported in the pre-clinical literature of psychotic disorders. Most items are reported in fewer than 30% of publications in the literature, with allocation concealment, the

performance of a sample size calculation and the availability of a study protocol most poorly reported among studies. The only item of study design criteria that showed high levels of prevalence for reporting in the literature, was the reporting of compliance with animal welfare regulations. This was reported in about 80% of studies published overall, and its prevalence of reporting has increased over time. Nevertheless, this number is still concerning considering that the legislation in modern societies demands that experimentation involving animals meets ethical guidelines (Kolar, 2006).

Unfortunately, this observation is not exclusive to the psychosis research field and seems to follow the trend of results from similar reviews in other areas of pre-clinical neuroscience. Data from fields of experimental focal cerebral ischaemia (Macleod et al., 2008), experimental autoimmune encephalomyelitis (EAE) (Vesterinen et al., 2010), pre-clinical Parkinson's disease (Rooke et al., 2011) and pre-clinical bone cancer-induced pain (Currie et al., 2013) show on average about 20% of papers report randomization of animals to groups and 26% report blinding of outcome measurement across these fields. Some of these reviews have found on average that less than 2% of publications report the performance of an *a priori* sample size calculation. In the psychosis research field, I have found this to be reported in less than 1% of publications. This is an important aspect of a study as the probability of detecting a difference of a certain size between a control and treatment group of animals is based on the number of animals that are used in each group, the size of the difference and the variability in the outcome (Macleod et al., 2009). A meta-analysis of EAE studies found that effect sizes reported were lower in studies using a larger number of animals in their experiments (Vesterinen et al., 2010).

Furthermore, it is recognised that in an experiment the reliability of a conclusion of a causal relationship between treatment and outcome is dependent on the internal validity of an experiment and its statistical power (Sena et al., 2014). This poor prevalence of reporting of important methodological criteria in the literature calls for concern as many of these measures have been shown to influence overall outcomes. It is important to assess for these methodological criteria, as research shows that animal studies that do not report randomisation or blinded assessment of outcome and perform experiments with no evidence of a sample size calculation, can give inflated effect sizes as opposed to studies that do report having carried out these measures (Sena et al., 2014). This can lead to an overstatement in both the

severity of a model and the efficacy of a drug tested. For example, in systematic reviews looking at the effects of hypothermia in experimental stroke, studies that did not randomize, overstated the reduction in infarct volume by 27% (van der Worp et al., 2007). Similar observations have been made in reviews of other fields (Currie et al., 2013; Macleod et al., 2008; Rooke et al., 2011).

From the data extracted in the context of publications describing developmentally-induced animal models of psychotic disorders, only the reporting of random allocation to group in treatment testing studies significantly affected outcomes reported. Studies that did not report randomising overstated the efficacy of treatment drugs tested when compared to studies that did report randomising. There were no significant differences in the effect of reporting or not reporting other risk of bias items in both model characterising and treatment testing studies. Based on these results, we might conclude that the reporting of risk of bias items does not have an effect on outcomes reported within pre-clinical psychosis publications. Or perhaps we might see a different pattern if the number of studies reviewed was larger and thus more representative of the entire field, not just one group of models. Of course, there is also the possibility that reporting of these measures does not necessarily reflect the active performance of these measures to reduce the risk of bias within the studies reviewed here. Therefore, there might be some interference from studies which report, but do not correctly perform these measures, or vice versa, studies that do perform these measures, just simply do not report them within their publications.

6.4.1 Change in reporting over time

In light of these observations in recent years, there has been an increasing amount of focus placed on communicating the importance of taking measures to reduce the risk of bias within experimental work using animals and improving transparency of the reporting of this research (Landis et al., 2012). This led to the publication of The Animal Research: Reporting In Vivo Experiments (ARRIVE) guidelines (Kilkenny et al., 2010a) and the Gold Standard Publication Checklist (GSPC) (Hooijmans et al., 2010a) in 2010. These guidelines took inspiration from established reporting guidelines for clinical trials (i.e. CONSORT Statement for randomised controlled clinical trials) and were written in a checklist format to provide guidance on how to report animal research. Since then, a number of journals have included the ARRIVE

guidelines in their guidance to authors when submitting research for publication (Baker et al., 2014).

In the preclinical literature describing animal models of psychotic disorders, there seems to have been an increase in the prevalence of reporting criteria, such as the reporting of compliance with animal welfare regulation and a statement of potential conflict of interest since the publication of the ARRIVE guidelines. However, while there has been an increase in the number of publications reporting measures taken to reduce these risks of bias, this improvement is offset by the increase in the number of studies published each year. Moreover, none of these guidance materials have led to an increase in the reporting of the performance of a sample size calculation. Therefore, we find that in the psychosis research field at least, there has been little change in the prevalence of the reporting of measures taken to reduce important risk of bias items such as blinded assessment of outcome and random allocation to group. In fact, the data show that these are two items that have been reported since the 1970s and the 1980s, respectively, and their prevalence of reporting in the field has varied little since then. This tells us that these concepts are not new in the field of animal studies and that despite the increase in awareness of the importance of reporting these measures within animal experiments, still very few studies do. This suggests that perhaps there are reasons, other than a lack of knowledge of the importance of taking these measures, to blame for this poor prevalence in reporting of these measures. It might be explained by the argument that just because something is not reported within a publication, doesn't mean it has not been done. It is true that we cannot ascertain the actual percentage of studies that have carried out certain study design measures, however, based on a review in experimental stroke, which found that there is little discrepancy between actual and reported levels of study quality (Samaranayake, 2006), this is unlikely to change overall results substantially. These unsatisfactory results in the prevalence of reporting of these study design measures might also be explained by the fact that while reporting guidelines are endorsed by many universities, journals and funding agencies, their completion is not mandatory for publication. This means that despite increased education on the matter and available forms of guidance, reporting standards have and are unlikely to improve in future unless further changes are made in the reporting process (Baker et al., 2014).

6.4.2 Differential reporting based on study design

Another reason behind poor reporting standards might be that they are influenced by the type of studies that are being performed and whether investigators think that blinding assessment of outcome for example is important or not for their experimental design.

When looking at whether different aspects of study design affected the reporting of methodological quality items, I observed that the inclusion of a conflict of interest statement for example was more prevalent among studies that tested the effect of a treatment drug in an animal model of psychosis compared to model characterising studies. This perhaps is a legacy of common practise in clinical trials and knowledge that financial ties of investigators positively influence the reported efficacy of drugs tested in support of these funders (Lundh et al., 2017).

When looking at how using different models within experiments describing in vivo experiments of psychotic disorders, many of the trends we see from the data can be explained by common sense. I noticed that for example in genetic models, the prevalence of random allocation to group was lower than in studies using other methods of model induction. For studies where genetic models are bred over many generations the reporting of randomisation of these models and wildtype models to "intervention" and "control" groups in model characterising studies hardly makes sense, as these animals are already different and are not being chosen from the same pool of animals. In contrast for lesion and developmental studies, where the prevalence of reporting of this risk of bias item were the highest, this is easily done as the independent variable is the location of the lesion or the difference in rearing environment that animals are subjected to and not the animal population itself. Interestingly however, results also showed that blinded assessment of outcome, for example, was most widely reported in studies reporting using genetic models. Arguably, for some experiments using these models, this is perhaps more difficult to do in these studies as opposed to studies using other models, due to obvious phenotypic differences between animals of different genetic backgrounds. In addition, it was interesting to note that pharmacological studies seemed to report important measures to reduce the risk of bias most poorly out of all studies, despite the fact that arguably it is easiest to implement these measures in these types of experiments.

Similarly, when looking at differences between studies measuring different outcomes in experiments describing in vivo experiments of psychotic disorders, many of the results are unsurprising. For example, random allocation to group was less likely to be reported in studies measuring electrophysiological outcomes compared to studies reporting behavioural and neurochemical outcomes. It is arguable that not allocating animals randomly to group has a more substantial impact on especially behavioural outcomes, as variation in behaviour will be much greater than variation seen in other outcomes like electrophysiological or anatomical outcomes among different animals.

Blinded assessment of outcome was most widely reported in studies measuring anatomical outcomes and least commonly reported in studies measuring electrophysiological outcomes. Electrophysiological outcomes are measured using electrical recording techniques and therefore, outcomes of these experiments are less prone to subjectivity over other measures such as anatomical or behavioural measures. This will likely become more and more relevant for studies measuring behavioural outcomes as new, automated tools start to replace human scoring of behaviours.

Overall, these differences between studies using different experimental designs are unlikely to explain the overall poor prevalence in reporting of study design criteria. They do however, provide an indication of where these problems are the greatest and where improvements in future might be made. It also highlights that some studies need to be evaluated on a different scale to the rest as their design does not fit and requires less of the classical methods of random allocation to group or blinded assessment of outcome perhaps.

6.4.3 Limitations

When reviewing the literature, it is only possible to analyse the reported quality of a study and there is no way of knowing the prevalence of the actual performance of some of these items within an experiment.

Therefore, it is possible that some studies took measures to reduce the risk of bias but did not report these. While the two qualities should be considered as separate concepts, they also overlap in many ways, as good reporting means that methodological quality of a study is easier to assess (Samuel et al., 2016). If we are only able to derive reporting quality from a scientific publication, this will affect at

how accurately we can derive information from the literature perhaps (Samuel et al., 2016). Therefore, while perhaps some might argue it is not appropriate to judge an experiment's quality in a publication based on the reported quality of said experiment, appraising exact methodological quality without complete reporting is difficult to do especially in the animal literature.

Alternatively, it may be that a failure to report measures to reduce the risk of bias is an indirect or surrogate measure of some other aspect of study design or conduct, which is responsible for the observed bias. Perhaps the lack of efficacy in blinded studies may be due to some other characteristic of those studies, such as the drug being tested, or the species used. There may also be other factors not looked at that could affect reporting of some methodological quality items in the literature, and thus confound the effects seen of variables assessed here. This could include things such as the nationality of authors and country where research was carried out in, where differences could arise perhaps from differing levels of emphasis based on the reporting of these measures in various countries, and institutions.

The journal of publication might also affect how studies report their experimental methods and what key things they report in their publications, depending on what a journal requests or deems acceptable for publication.

Identifying these differences would be important, because under those circumstances addressing the issue of blinding would not address the underlying bias.

This also introduces the idea of covariance and that perhaps studies that do not report blinded assessment of outcome also don't report randomising animals to groups (van der Worp et al., 2010).

7 Capturing all the Data of Relevance

Systematic reviews can be extremely powerful in summarising a field quantitatively and highlighting areas where research may be lacking, or improvements can be made. These reviews and our knowledge based on them are strongly limited by the depth and breadth of research they are able to capture. Therefore, a review will be biased towards evidence that is made available in the literature and if this evidence is substantially different from that which is unpublished, this can affect conclusions drawn in these reviews. While systematic methods used aim to identify all sources of evidence out there, this step is still limited by the search strategy and terms used to identify relevant publications of data and how closely these reflect actual terms used in the literature.

In this chapter, I explore the robustness of my systematic search of the literature and assess the extent to which these limitations might have affected my overall findings. I estimate the amount of publication bias that could be present in the literature describing animal models of psychotic disorders, look at the effect of updating a search a year after the original search and explore the impact of performing the search again, but this time using alternative keywords to answer the same research question.

7.1 Missing data due to publication and reporting biases

It is well established that positive and promising studies are more likely to be published in the literature than negative or neutral studies and naturally this skews the resulting conclusions drawn about biological truth represented by the literature (Rosenthal, 1979). This can lead to the wrongful estimation of the overall efficacy of a treatment drug or the effectiveness of a model in recapitulating a disorder. In clinical trials, this issue has long been recognised (Easterbrook et al., 1991) and has led to the introduction of registration systems for clinical trials allowing for more transparency in clinical research (De Angelis et al., 2004). In recent years, a number of studies have shown that the pre-clinical research field is not exempt from this problem either (Sena *et al.*, 2010; ter Riet *et al.*, 2012), and there has been a lot of discussion around the use of similar registries for pre-clinical studies (Wieschowski et al., 2016).

Reporting or not reporting of evidence depending on certain characteristics such as the directionality of findings also happens in the form of selective outcome reporting bias. This occurs when non-significant outcomes are omitted from the publication of a study despite having been evaluated in the study (Ioannidis et al., 2014b).

Evidence for this concept is supported by observations that key findings cannot be replicated in many pre-clinical research fields (Begley and Ellis, 2012).

Nevertheless, it is difficult to estimate the exact extent of this problem as few pre-clinical studies publish protocols of their studies (as shown in Chapter 6). During categorization of included studies in this review, it became clear that many publications only reported data for outcome measures of significance or positive results. This selective reporting bias was very transparent in the sense that most studies reported having carried out the relevant behavioural measures, they just failed to report the actual data collected.

7.1.1 Methods

7.1.1.1 Source Data

As publication bias requires a large amount of data, this concept was explored using data extracted and analysed in Chapter 4. This data described 974 model-characterising experiments from 84 studies, and 143 treatment-testing experiments within 17 studies. I also briefly estimated the extent of selective outcome reporting bias within all identified studies in this systematic review.

7.1.1.2 Analysis

Methods used for meta-analysis are described in detail in Chapter 3. All data used for publication bias analyses were unnested and model-characterising comparisons were analysed separately to treatment-testing comparisons. Sensitivity analyses were performed by removing 5% of the most extreme data points and reanalysing results. For estimation of the extent of selective outcome reporting bias within published literature, all identified studies that could be analysed through text mining were processed for the phrase “data not shown” using the program *PDF-Xchange Editor*.

7.1.2 Results

7.1.2.1 Prevalence of publication bias in a subset of the psychosis research field

Visual inspection of funnel plots for model-characterising experiments and treatment-testing experiments suggested that there might be some evidence of publication bias for model-characterising experiments, but the graph was less clear for treatment-testing experiments (Figure 7.1).

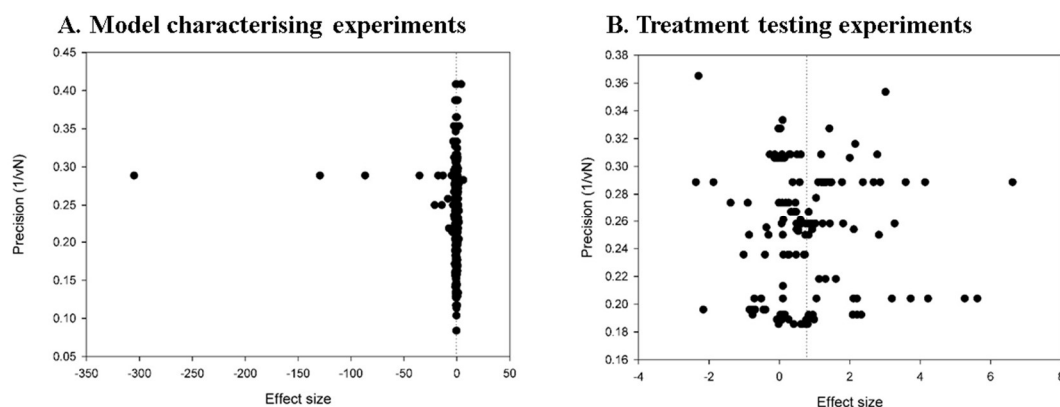


Figure 7.1 Funnel plot of model-characterising and treatment-testing experiments
Funnel plots showing precision plotted against effect size. In the absence of publication bias these plots should be symmetrical around the global effect size (i.e. the dotted line).

Egger regression suggested that there was significant asymmetry of model-characterising experiments, but not drug testing experiments (Table 7.1).

Table 7.1 Overall prevalence and potential impact of publication bias

	Reported overall effect size (95% CI)	Bias with Egger regression (P-value)	Bias with METATRIM	Number of additional studies considered missing	METATRIM adjusted effect size (95% CI)	Absolute difference in effect size
All model characterising experiments	-1.126 SD (-1.635 - -0.617)	+ (P<0.038)	+	458	-2.236 SD (-2.594 - -1.879)	1.11
Sensitivity analysis of model characterising	-0.472 SD (-0.506 - -0.437)	+ (P<0.019)	-	0	-	-
All treatment testing experiments	0.780 SD (0.552 - 1.008)	-	-	0	-	-
Sensitivity analysis of treatment testing experiments	0.697 SD (0.538 - 0.856)	-	-	0	-	-

Asymmetry was more apparent in model-characterising experiments using Egger regression than for treatment-testing experiments (Figure 7.2).

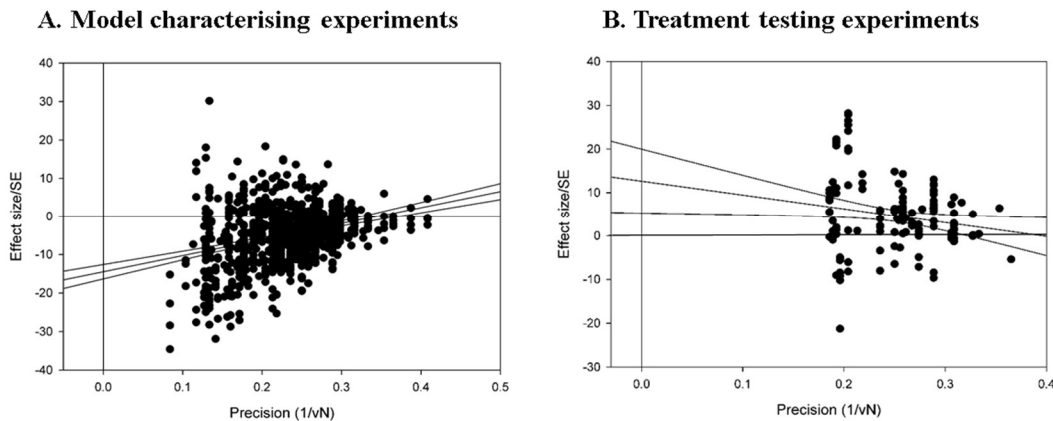


Figure 7.2 Egger regression plot of model-characterising and treatment-testing experiments
Egger regression showing precision against standardized effect size. Publication bias is measured by how close to the origin the intercept of the regression line is.

Finally, trim-and-fill analysis in STATA confirmed results of Egger regression – that treatment-testing studies showed no significant asymmetry and that the model-characterising dataset did show asymmetry.

7.1.2.2 Impact of publication bias in the field

Using trim-and-fill 458 experiments were estimated to be “missing” from the model-characterising dataset, suggesting that these studies were conducted, but not reported in the literature. Trim-and-fill also imputed an overall estimate of effect size taking into account these potentially missing studies, giving an estimate of -2.236 SD units (95% CI -2.594 - -1.879). This was 1.11 SD units below the original overall reported effect size and therefore suggested that the original reported effect size was an underestimation of the true effect. No studies were estimated to be missing in the treatment-testing dataset.

7.1.2.3 Sensitivity analysis of results

Due to obviously large outliers in the model-characterising dataset sensitivity analyses were run on both datasets. This 5% truncated mean analysis revealed no further evidence of potential publication bias for treatment-testing experiments,

however, overall reported effect size did reduce by 0.083 SD units (see Table 7.1 above). Sensitivity analysis did however, affect earlier estimation of publication bias in model-characterising experiments. Overall, reported effect size was reduced from -1.126 SD units to -0.472 SD units and while Egger regression still suggested there was potential evidence for publication bias, trim-and-fill analysis did not suggest any asymmetry affecting the dataset.

7.1.2.4 Selective outcome reporting bias

Altogether, PDF documents of 5622 publications identified in my original search could be searched through the program *PDF-Xchange Editor*. The program identified 8579 entries of the phrase “data not shown” in 3806 documents (68% of publications searched).

7.2 Missing data due to limitations across time and space

Systematic reviews are used to provide current, up-to-date and comprehensive evidence on a particular research question. A major limitation of bigger reviews especially is the time that it takes to complete them. Another issue is the ability to capture all that is relevant when you are asking a broad research question about an otherwise uniquely human disorder. The totality and relevance of the evidence that is gathered and drawn conclusions from in a systematic review, is based almost entirely on the words and phrases used to search the literature to obtain this evidence in the first place. I wanted to explore the possible extent of both of these limitations and the impact these might have had on my overall conclusions using a model of substance-induced psychosis.

7.2.1 Cocaine: a model of substance-induced psychosis

We have seen that pharmacological models have continued to predominate the field of animal modelling of psychotic disorders and many of these substances are drugs of abuse in human populations. Drugs with psychotomimetic properties such as cannabis, cocaine, phencyclidine and amphetamines are able to produce psychotic symptoms similar to those seen in schizophrenia (Connell, 1990; Steeds et al., 2015). Intoxication from substances can cause acute psychotic effects, but also a chronic abuse of these substances can lead an increased risk of developing more substantial psychotic outcomes which are independent of these temporary effects of drugs (Moore et al., 2007). Stimulants given to schizophrenic individuals can also

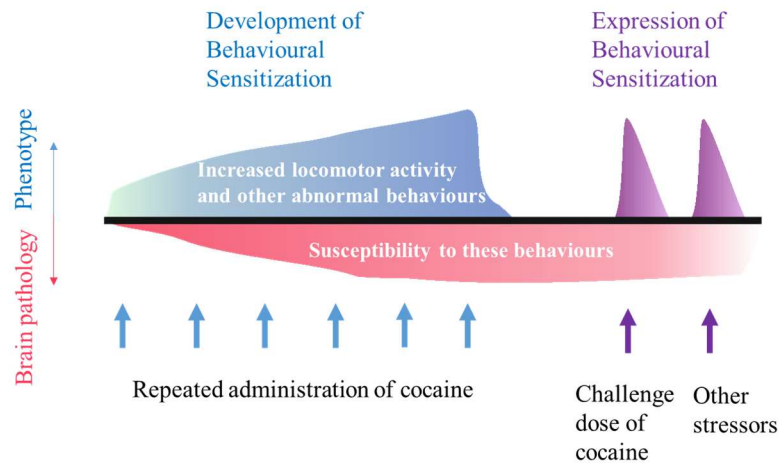
provoke a psychotic state which is almost identical to their own positive symptoms (Janowsky et al., 1973). According to the DSM-5, substance-induced psychotic disorders are characterized in the clinic by prominent delusions and/or hallucinations that an individual experiences during or shortly after substance intoxication and these symptoms can persist for weeks. As stimulant-induced psychotic symptoms usually subside within several days to a month after the end of substance abuse, the persistence of symptoms for more substantial periods of time might be better explained by a primary psychotic disorder (American Psychiatric Association, 2013; Sachdev and Keshavan, 2010).

Cocaine is a substance of abuse, taken by about 18.2 million people throughout the world (United Nations Office on Drugs and Crime, 2016). In Europe, it is estimated that 4.1% of the population between the ages 15 and 64 have used cocaine at least once during their lifetime (European Monitoring Centre for Drugs and Drug Addiction, 2010). Its use is linked to psychiatric problems, with psychotic symptoms being most typical (Vergara-Moragues et al., 2014). Chronic use of the substance induces a paranoid psychotic state directly related to drug use and is almost indistinguishable from symptoms seen in acute paranoid schizophrenia (Brady et al., 1991). It is thought that like with many other psychostimulants, once use of cocaine stops, hallucinations usually also end, but delusions can linger longer.

Cocaine is an indirect dopamine agonist and in animals induces hyperlocomotion, consisting of ambulation and seeking behaviour. This is thought to occur through sensitization of the dopamine system (Ujike, 2002). Chronic administration experimental paradigms where animals are given repeated administrations of a substance induces a phenomenon called 'behavioural sensitization', which has two distinct stages – 'development' and 'expression' (Richtand, 2006; Ujike, 2002). Sensitization is defined as a non-associate learning process where repeated exposures to a stimulus result in gradual augmentation of the behavioural effect (Weidenauer et al., 2017). This means that in these experimental paradigms locomotor activity gradually increases with repeated administration of the drug, a phase called the development of sensitization (Figure 7.3). This sensitized state also persists for a long time, evidence showing that in rats it remains a year after abstinence from substances such as amphetamine (Paulson et al., 1991). Once sensitization develops, challenging the animals after a period of abstinence and withdrawal from the drug with any subsequent doses of the drug will produce

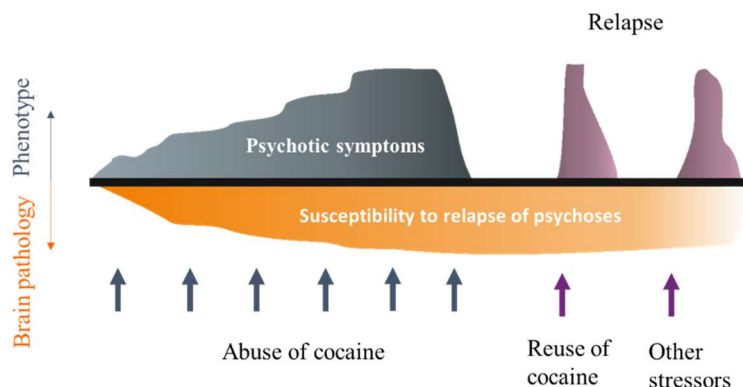
intense stereotypy (Ujike, 2002). This is called the expression stage of sensitization. This enhanced activity is also observed in response to other stimulants as well as other types of drugs such as morphine, nicotine and cannabis and environmental or physiologic stressors (Ujike, 2002). The two phases – development and expression – of behavioural sensitization are thought to be different anatomically and neurochemically (Reeves et al., 2004).

Figure 7.3 Effects of chronic administration of a stimulant like cocaine to animals in the laboratory (Figure adapted from Ujike, 2002).



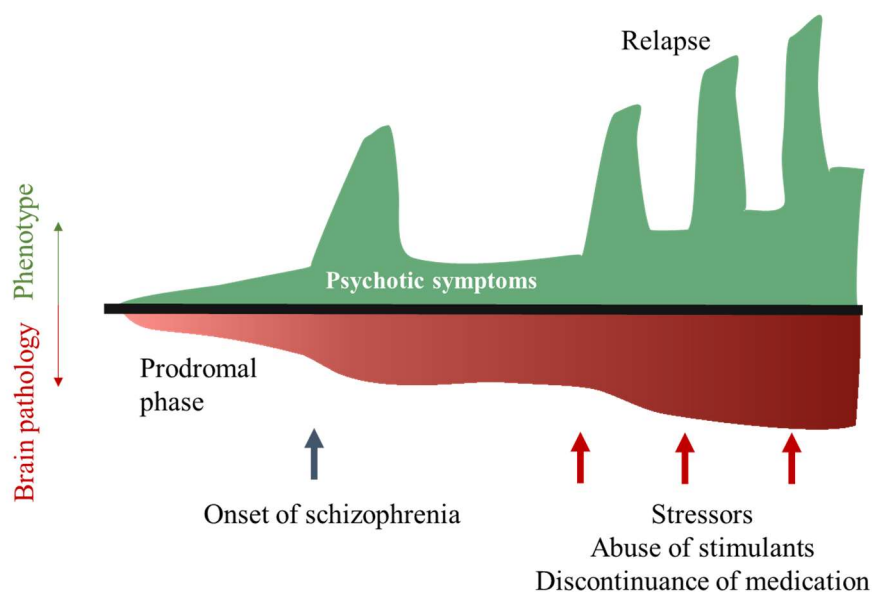
This sensitization also occurs in humans, where chronic, intermittent use of psychostimulants induces psychosis through the sensitization of dopaminergic systems (Figure 7.4). It is thought that initially chronic abuse of cocaine creates symptoms of euphoria, followed by dysphoria, and finally paranoid psychosis, increasing in severity with increasing dose of cocaine and increasing chronicity of cocaine use (Post, 1975). While withdrawing from the drug of abuse usually resolves psychotic symptoms within a short period of time, this psychotic state, resembling that of the initial symptoms can be induced again with few, or even a single exposure to the substance at lower doses than before (Sato et al., 1983). In addition, psychologic stressors and other drugs of abuse can have the same effect and these effects are seen even years or decades after abstinence (Ujike, 2002).

Figure 7.4 Effects of chronic abuse of cocaine in humans (Figure adapted from Ujike, 2002).



Chronic paranoid schizophrenia is thought to share these mechanisms of underlying behavioural sensitization (Figure 7.5). Some individuals with schizophrenia exhibit activation and exacerbation of psychotic symptoms as a result of acute exposure to psychostimulants at doses otherwise not psychotogenic in healthy individuals (Curran et al., 2009; Lieberman et al., 1987). PET studies also show that dopamine release is increased in response to amphetamine, showing an exaggerated response in schizophrenic individuals (Breier et al., 1997). Since stimulant-induced increase in dopamine release characterizes sensitization, this has led to the concept of schizophrenia possibly representing a state of endogenous sensitization (Laruelle, 2000). Evidence shows that increased dopamine activity already occurs at prodromal stages of the disorder, which predates the onset of psychotic symptoms (Howes and Kapur, 2009). It is thought that here and during relapses, dopaminergic neurons are hyperresponsive to outside stressors such as environmental stimuli, stimulant use or discontinuance of medication (Laruelle, 2000; Ujike, 2002). This means that the pathologic process is thought to have already begun during the prodromal phase, and the disease and full-blown psychotic symptoms become fully manifested when the threshold is exceeded (Ujike, 2002).

Figure 7.5 The sensitization model of schizophrenia (Figure adapted from Ujike, 2002).



In the clinic, psychotic symptoms presenting with cocaine abuse can be diagnosed as *cocaine intoxication*, which is more of an acute state and disappears with abstinence; *psychotic disorder induced by cocaine*, where psychotic symptoms last longer and are usually more severe; or *schizophrenia* with cocaine abuse, where

individuals would be diagnosed with the primary psychotic disorder but can experience acute psychotic symptoms induced by cocaine at the same time. These diagnoses mainly differ on the duration of their symptoms and therefore it is often difficult to give a distinguished diagnosis straight away (American Psychiatric Association, 2013; Vergara-Moragues et al., 2014). In the clinic, all three of these diagnoses are treated using the same pool of antipsychotic drugs. These dopamine receptor blocking agents have been shown to be able to block the development of sensitization when given during the induction of this sensitization and suppress the expression of sensitized behaviour when given before a psychostimulant challenge, however, have not been seen to be as effective in reversing the sensitized state (Meng et al., 1998; Shuto and Nishi, 2011). Therapeutic agents that are able to reverse this state of sensitization instead of just controlling it may have potential therapeutic value for schizophrenia. Moreover, as substance abuse in schizophrenic individuals is common and can cause a relapse or worsening of symptoms, better treatment options are required for individuals with co-morbidities (Curran et al., 2009).

7.2.2 Are we lagging behind in light of new data: the impact of updating a search

My aim was to explore the extent to which updating my original search would change overall results.

7.2.2.1 Methods

This work was carried out as part of a BSc Biological Sciences Honours dissertation project and I would like to acknowledge Fala Cramond (FC) for her work on performing the update search, screening of studies for inclusion and exclusion and extracting data for included publications. Data were checked and meta-analyses were run by myself.

7.2.2.2 Search strategy

3847 publications identified to be of relevance in the current review as described in Chapter 3, were filtered by searching for “cocaine” in the title, abstract and keywords. This gave a subset of the data that will be hereon referred to in this chapter as the “Original search”. In addition to this, an update search of PubMed was completed in January 2015 using the same string of search terms as described

in Chapter 2 and restricted to studies published after January 2014. The same filter was applied to this dataset as the original dataset to give a subset of studies exploring the effects of cocaine in animals, labelled from hereon as the “Update Search”. Publications were screened independently by title and abstract in MS Access by two reviewers (FC and AS) and discrepancies were resolved by a third screener (myself).

7.2.2.3 Inclusion criteria

Initial methods used for screening were identical to those described in Chapter 2, with the addition of the following criteria: only experiments describing the effects of cocaine on locomotion were included. Studies specifically investigating the addictive effects of cocaine, those testing cocaine in transgenic mice and those that used animals with co-morbidities were also excluded. Experimental paradigms using both acute and chronic administration of cocaine were included, and this was not restricted to any particular route of delivery. The only outcome measure was horizontal locomotion and therefore other outcome measures, such as stereotyped behaviour, were not extracted.

7.2.2.4 Data extraction

From each study included, experimental comparisons describing the effects of acute or repeated administration of cocaine on locomotion and those exploring modulation of this effect through the administration of therapeutic agents were included. These comparisons were extracted and analysed separately. A single reviewer (FC) extracted study design, quality and outcome data for each included comparison as described in Chapter 2. For experiments describing acute effects of cocaine, where data were presented as total locomotion over a period, total movement and time period of assessment was extracted. Where locomotion was reported over a period through activity at multiple time points, the mean activity over that time was taken and time period of assessment was recorded as final time point of assessment. In experiments describing chronic cocaine administration paradigms, where the development of behavioural sensitization was measured by reporting locomotor activity on the first and last day of cocaine administration, the difference between these two measures was taken. For experiments measuring the expression of behavioural sensitization by reporting locomotor activity in response to a challenge dose of cocaine after withdrawal from the repeated administration paradigm, the first

withdrawal time point was taken. 15% of all data extracted were checked by me and any errors that were encountered within this set were changed. As no consistent errors were identified within this subset of the data, everything was taken ahead for further analysis. Overall, both raw data and analyses were also subject to a sense check and any data that appeared erroneous was double-checked within the publication.

7.2.2.5 Analysis

Meta-analysis methods are as described in Chapter 2. Acute cocaine-induced locomotor activity and chronic cocaine-induced sensitized locomotor activity were analysed separately. Moreover, for chronic cocaine administration paradigms, experiments measuring the development of behavioural sensitization and later measurements of expression of behavioural sensitization in response to a challenge dose of cocaine were also analysed separately. Where appropriate and data were sufficient (i.e. over 25 comparisons included in the meta-analysis), univariate meta-regression was performed to investigate potential sources of heterogeneity. Meta-regression was used to investigate possible sources of heterogeneity including components of study quality and methodological criteria checklist and study design characteristics, where a significance level of $p < 0.05$ was set for each test. To correct for multiplicity of testing a Holm-Bonferroni adjusted critical p value was calculated to account for the number of variables tested within subgroup analyses, calculated separately for study quality and study design items. For study quality items and for study design items explored in experiments using chronic cocaine administration paradigms adjusted critical p value was set at $p < 0.006$. For study design items explored in experiments using acute cocaine administration paradigms and characterising the model adjusted critical p value was set at $p < 0.009$ and for those looking at the effect of a treatment on models of cocaine-induced psychosis it was set at $p < 0.007$.

This included variables of study quality and study design characteristics. Criteria for study quality were as specified in Chapter 2. Study design characteristics for all studies were also as specified in Chapter 2, including (1) species of animal used, (2) strain of animal used, (3) sex of animal used, (4) dose of cocaine administered (i.e. whether acutely or repeatedly), (5) route of cocaine administration, (6) time of outcome assessment – measured as duration of assessment for acute locomotor activity and as last day of cocaine-induced locomotor activity assessment for chronic cocaine experiments. Furthermore, for experiments measuring locomotor activity after chronic exposures to cocaine, (7) number of cocaine administrations was tested, and for those measuring behaviour after withdrawal from this experimental paradigm (8) days of withdrawal between last cocaine administration and challenge dose of cocaine administration were also tested (Figure 7.6). For experiments testing the effect of a treatment drug (9) time of treatment administration and (10) treatment administered were also explored.

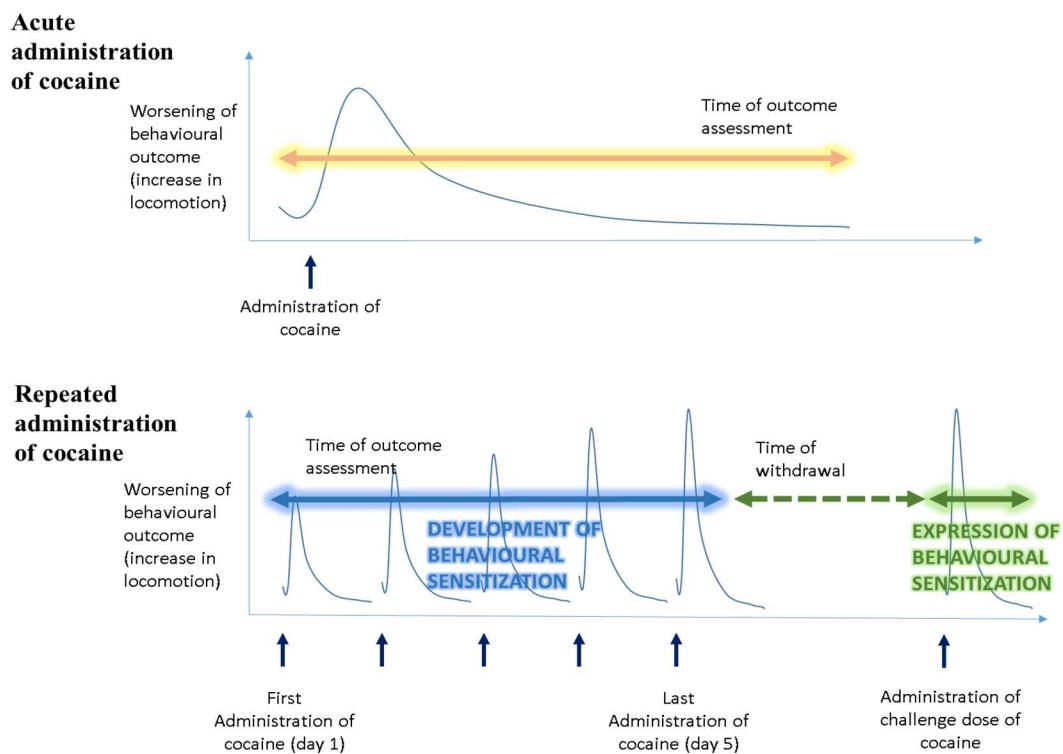


Figure 7.6 Experimental design of acute and chronic administration paradigms of cocaine induced psychosis in animal models

7.2.2.6 Results

457 of the initially identified 14,721 publications identified in my main search included the word “cocaine” in their title, abstract or keywords. 153 were found to be potentially relevant to our research question at this stage and included. At full-text screening 61 publications of relevance were identified describing both model-characterising and treatment-testing experiments. Updating the search a year later added another 12 publications of relevance.

Overall, measures to reduce bias were reported in few publications (Table 7.2). In the original dataset the median number of study quality checklist items scored was one (interquartile range [IQR] 1-2). Random allocation of animals to group was reported in 5 publications (24.2%), blinded assessment of outcome in 6 (9.7%), reporting of animals excluded from analysis in 5 (8.1%), a statement of potential conflict of interest in 9 (14.5%), and compliance with animal welfare regulations in 40 (64.5%). No publications reported having carried out a sample size calculation, the availability of a protocol for their study or having blinded the induction of model or administration of treatment. Improvements were only seen in the reporting of compliance with animal welfare regulations (increased to 70.3%) and potential conflict of interest statements (20.3%) when the search was updated to include the additional studies identified a year after the original search. In contrast to the original search, no studies in the updated search reported blinded assessment of outcome or animal exclusions.

Table 7.2 A comparison of the reporting of methodological quality items and measures taken to reduce risks of bias in studies identified in our original search, in the updated search and with all the data pooled

	Original Search	Updated Search	Overall
Total No. of publications identified	61	12	73
Number of publications reporting ... (Overall %)			
Random allocation to group	15 (25%)	2 (16.7%)	17 (23%)
Allocation concealment	0	0	0
Blinded assessment of outcome	6 (10%)	0	6 (8%)
Sample size calculation	0	0	0
Compliance with Animal Welfare Regulations	40 (66%)	12 (100%)	52 (71%)
Statement of potential conflict of interest	9 (15%)	6 (50%)	15 (21%)
Exclusion of animals pre-specified or explained	5 (8 %)	0	5 (7%)
Availability of a study protocol	0	0	0
Median quality (/8) (interquartile range)	1 (1-2)	2 (1-2)	1 (1-2)

7.2.2.6.1 Acute cocaine-induced hyperactivity

7.2.2.6.1.1 Results of the original search

Acute locomotor activity was reported in 51 publications in the original dataset and 9 in the updated dataset. Due to the small number of studies identified in the updated search, the data were insufficient to perform meta-regression on, but they were incorporated into an overall model that included both searches totalling data from 60 publications (Table 7.3).

Table 7.3 A comparison of the data extracted from studies reporting acute cocaine-induced locomotor activity in the original search and in the updated search
Negative values indicate a worsening in outcome and positive values indicate an improvement in the behavioural outcome of the animal.

Acute locomotor activity		Original Search	Overall
No. of publications		51	60
Animals/paper		46.3	43.69
Naïve vs. Model Animals	Effect size	-1.0689 SD	-1.063 SD
	(95% CI)	(-1.2688 - -0.8689)	(-1.2496 - -0.8765)
	I^2 (%)	46.8	49
	No. of experiments	66	81
	No. of animals	1234	1487
Model vs. Treated Animals	Effect size	0.9293	0.9185
	(95% CI)	(0.6958 - 1.1629)	(0.6912 - 1.1457)
	I^2 (%)	52.2	51.2
	No. of experiments	91	93
	No. of animals	1134	1163

Overall, 66 experiments testing the effects of cocaine on locomotion (using 1234 animals) were identified and extracted. 8 of these experiments were performed in gerbils (243 animals), 20 using mice (338 animals) and 38 using rats (653 animals). The most commonly used strain was Sprague Dawley (29 experiments, 566 animals, representing 43% of all animals used and 87% of rats). There was much less consistency in the strain of mouse used, the most common being C57Bl/6 mice (6 experiments, 102 animals, representing 30% of mice). Most experiments used male animals (48 experiments, 875 animals), while 5 (131 animals) used female animals, 4 (51 animals) used both and 9 (177 animals) did not state the sex of animal used. Dose of cocaine used in the experiments varied from 0.04 mg/kg to 56 mg/kg, with 58 experiments (88%) using 20mg/kg or less. Time of behaviour assessment ranged from 10 minutes to 3 hours.

Altogether, administration of cocaine increased locomotor activity by 1.0689 SD units (95% CI 0.8689 to 1.2688). This increase in locomotor activity is defined as a worsening in behavioural outcome, and therefore an increased efficacy of the model at showing signs of hyperactivity thought to model the psychotic symptoms seen in humans. The dataset also showed moderate heterogeneity ($\tau^2 = 0.292$, $I^2 = 46.8\%$), which was only significantly explained by the dose of cocaine used to induce the model ($p = 0.0031$, $\tau^2 = 0.2201$, $I^2 = 41.88\%$, $\text{adj } R^2 = 24.62\%$, Figure 7.7 Relationship between dose of cocaine used to induce the model and reported effect size in model-characterising studies (Figure 7.7). Heterogeneity was not explained by any of the other study characteristic variables investigated with univariate meta-regression, namely: animal, strain and sex of animals, dose of cocaine used to induce the model, method of cocaine administration and time of outcome assessment.

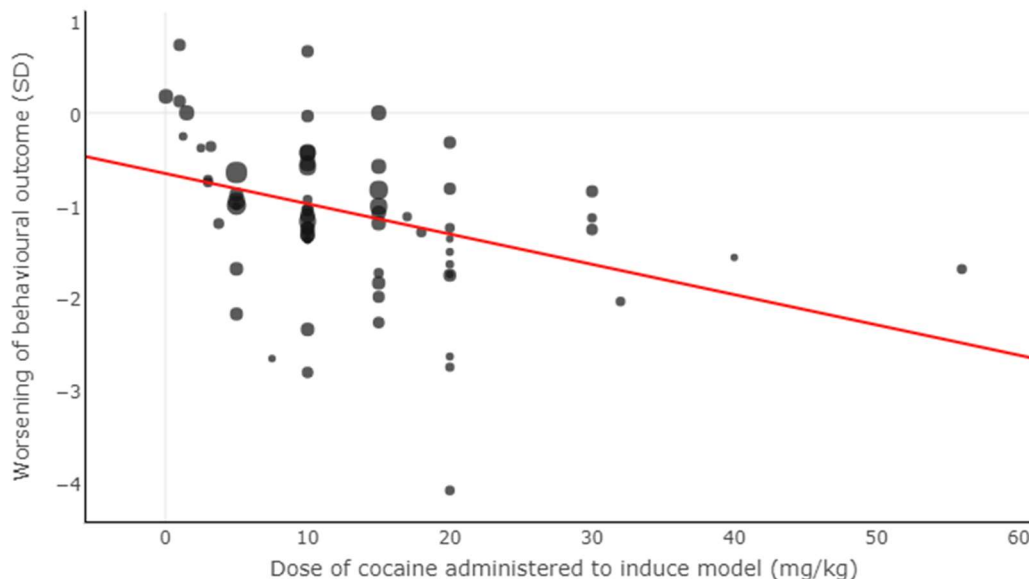


Figure 7.7 Relationship between dose of cocaine used to induce the model and reported effect size in model-characterising studies

For study quality, a greater increase in locomotion was observed and therefore worse behavioural outcome in studies that reported random allocation to group as opposed to studies that had not reported taking measures to reduce this risk of bias ($p = 0.0053$, $\tau^2 = 0.2131$, $I^2 = 40.94\%$, $\text{adj } R^2 = 27.04\%$, Figure 7.8). No other potential sources of bias or other methodological quality accounted for a significant proportion of heterogeneity.

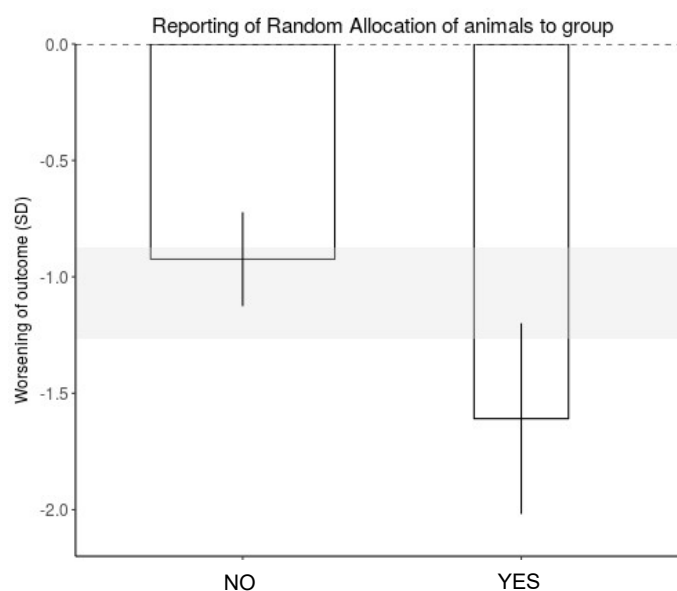


Figure 7.8 Relationship between the reporting or not reporting of randomisation and reported effect size in model-characterising studies

In this original search, an additional 91 experiments (using 1134 animals) were identified as testing the efficacy of interventions in moderating cocaine-induced hyperlocomotion.

15 of these experiments were performed in gerbils (328 animals), 29 using mice (256 animals) and 47 using rats (550 animals). The most commonly used strains were Sprague Dawley (24 experiments, 328 animals, representing 29% of all animals used and 60% of rats) and Mongolian gerbils (15 experiments, 328 animals, representing 29% of all animals used and all of the gerbils used). The most common strain of mouse used was Swiss-Webster mice (15 experiments, 126 animals, representing 49% of mice). An overwhelming majority of experiments used male animals (88 experiments, 1094 animals), while the rest of the experiments did not state the sex of animal used (3 experiments, 40 animals). Dose of cocaine used in the experiments varied from 0.2 mg/kg to 40 mg/kg, with 89 experiments (99%) using 30mg/kg or less. Time of behaviour assessment ranged from 10 minutes to 2 hours. Time of treatment administration varied from three days prior to and up to half an hour after cocaine administration.

Considered together, drug treatments improved hyperlocomotion induced by acute cocaine by 0.929 SD units (95% CI 0.6958-1.1629), with again moderate

heterogeneity ($\tau^2 = 0.5014$, $I^2 = 52.2\%$). The use of different drugs explained part of the heterogeneity observed ($p = 0.0009$, $\tau^2 = 0.2071$, $I^2 = 28.79\%$, $\text{adj } R^2 = 58.07\%$, Figure 3), however, many of these were only tested in one or two experiments. Nevertheless, I found that the two most effective drugs in reducing hyperlocomotion were amperozide, an atypical- (5.04 SD units, 95% CI 2.7673 - 7.3105) and fluphenazine (5.00 SD units, 95% CI 2.9902 - 7.0093), a typical-antipsychotic. Drugs grouped by their mechanism of action did not contribute significantly to heterogeneity.

7.2.2.6.1.2 Results from updating the search

Updating the search identified an additional 17 experiments using 281 animals. 15 experiments described model-characterising paradigms and 2 experiments looked at the effect of a treatment administered to modulate cocaine-induced locomotion. Both treatment exploring experiments were from the same study and used male, Sprague Dawley rats. Model-characterising studies used mice (7 experiments, 129 animals) and rats (8 experiments, 124 animals) in about equal proportion. All rats were Sprague Dawley, while the most common strain of mice used was C57BL/6 (4 experiments, 81 animals). As seen before the majority of experiments reported using male animals (9 experiments, 125 animals), while one experiment reported using females (14 animals) and five experiments did not specify the sex used (114 animals). The dose of cocaine administered among different experiments to induce the model varied from 0.15 mg/kg to 20 mg/kg and time of outcome assessment ranged from 15 minutes to 3 hours.

Both of these datasets were too small to perform univariate meta-regression on, but data were combined with those from original search to give an overall, pooled dataset. By pooling data from original search and those identified in the following year in the literature, effect sizes changed little. Administration of cocaine increased locomotor activity by 1.063 SD units (95% CI 0.8765 to 1.2496), with still moderate, albeit slightly more heterogeneity than seen in original dataset alone ($\tau^2 = 0.32$, $I^2 = 49.0\%$). As before, part of the heterogeneity was significantly explained by dose of

cocaine administered ($p = 0.0001$, $\tau^2 = 0.2113$, $I^2 = 40.37\%$, $\text{adj } R^2 = 33.96\%$, Figure 7.9), but not by any other variables assessed.

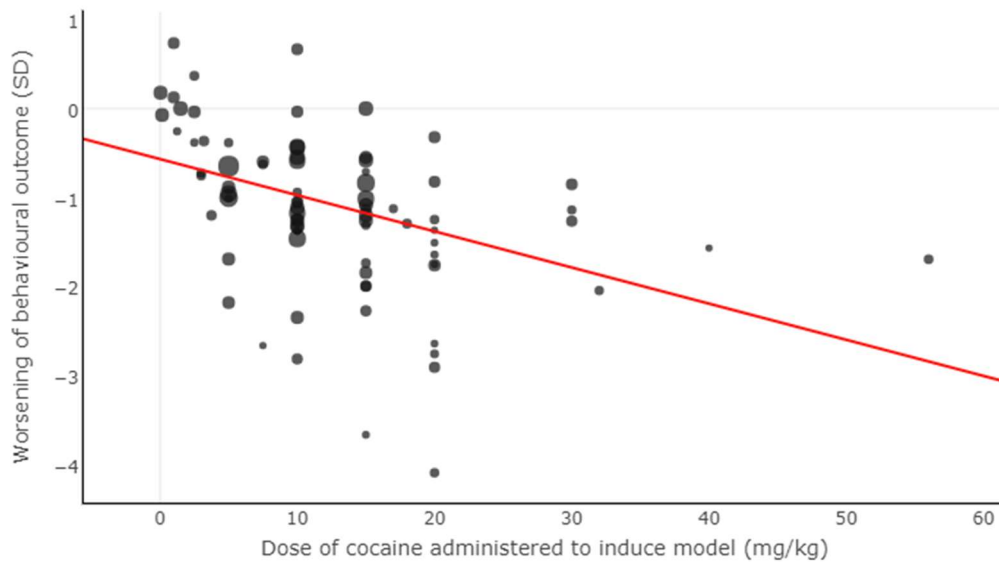


Figure 7.9 Relationship between the dose of cocaine used to induce model and reported effect size in model-characterising studies from the alternate search

Overall, drug treatments improved hyperlocomotion induced by cocaine by 0.919 SD units (95% CI 0.6912-1.1457), with again moderate, however, slightly less heterogeneity than seen in just the original dataset ($\tau^2 = 0.4722$, $I^2 = 51.2\%$). This heterogeneity was like before, in part, only significantly explained by the type of drug administered as treatment ($p = 0.0009$, $\tau^2 = 0.1969$, $I^2 = 27.80\%$, $\text{adj } R^2 = 58.30\%$). The most effective drugs remained the same as before. No other variables of study characteristics or quality contributed significantly to heterogeneity.

7.2.2.6.2 Chronic cocaine-induced locomotor activity

Experimental paradigms involving a chronic drug administration schedule were identified in 22 publications through the original search. Both model-characterising and treatment exploring experiments using these paradigms measured either the development of sensitization over a period of time during which the animal was administered cocaine, as well as the expression of sensitization measured after abstinence from cocaine for a set period of time. These were analysed separately as we recognise them to be different in terms of measuring underlying biology. Individually, none of these separate meta-analyses were sufficient for univariate meta-regression, therefore heterogeneity could not be explored in detail.

From these 22 publications, 56 experiments using a repeated cocaine administration paradigm were identified. 18 reported measuring the development of behavioural sensitization. I define the development of behavioural sensitization here as the difference in locomotor activity induced by cocaine given on the first day of drug exposure compared to that induced by cocaine on the last day of the repeated drug administration paradigm. In addition, 38 experiments reported measuring the expression of sensitization. This is defined as an increase in locomotor activity in response to administration of a challenge dose of cocaine after a certain period of withdrawal from the original chronic drug abuse paradigm.

7.2.2.6.2.1 Experiments characterising the model

7.2.2.6.2.1.1 Results of the original search

Out of the total 56 experiments, 10 experiments (using 208 animals), measured the behavioural effects of repeated cocaine administration on the development of behavioural sensitization. In addition, 14 experiments (using 283 animals) measured the expression of sensitization produced after a period of abstinence and a subsequent administration of a challenge dose of cocaine. Of these model-characterising experiments, species used were mainly mice (15 experiments, 345 animals), of the strains C57BL/6J (4 experiments, 92 animals) and Swiss-Webster (4 experiments, 100 animals), but also Balb/c (2 experiments, 46 animals), ICR (3 experiments, 76 animals), CD-1 (1 experiment, 11 animals) and Swiss-Albino (1 experiment, 20 animals). The only other species used were rats (8 experiments, animals), including strains Sprague Dawley (6 experiments, 123 animals), Lewis (1 experiment, 9 animals) and Wistar (1 experiment, 14 animals). More of these animals used were male (14 experiments, 288 animals, 59% of all animals used), than females (1 experiment, 32 animals), while only 3 experiments used both (60 animals) and 5 did not state gender of their experimental subjects (111 animals). Dose of cocaine administered repeatedly was either 7.5mg/kg (1 experiment), 10 mg/kg (12 experiments), 15mg/kg (8 experiments) or 20 mg/kg (2 experiments). These repeated injections were mostly administered once a day, or twice a day for 1 experiment, for the duration of either 5 (16 experiments), 7 (2 experiment), 8 (4 experiments) or 20 (1 experiment) days, repeatedly. For those experiments measuring it, time of withdrawal allowed between the development of sensitization and measurement of expression of sensitization ranged from 2 days to 28 days after abstinence to cocaine.

Altogether, administration of cocaine over a prolonged period amplified locomotor activity by 3.4024 SD units (95% CI 2.2054 to 4.5993,

Table 7.4) on the final day of administration compared to the first day of administration. This augmentation in locomotor activity is thought to indicate a development of sensitization to cocaine in these animals. Substantial heterogeneity was observed in the data ($\tau^2 = 2.2523$, $I^2 = 83.5\%$), however, I could not explore this further using univariate meta-regression due to the small sample of data. When animals were tested after withdrawal from the chronic cocaine administration paradigm, a cocaine challenge increased locomotor activity by 1.3757 SD units (95% CI 0.7736 – 1.9778). This is thought to be a measure of cocaine-induced expression of behavioural sensitization, or in other words the effect of chronic cocaine abuse on the animals' susceptibility to hyperlocomotion in response to subsequent uses of the substance.

Table 7.4 A comparison of the data extracted from studies characterising chronic cocaine-induced animal models in the original search and in the updated search.

Chronic locomotor activity		Original Search	Overall with Updated Search
Total no. of publications		22	31
Animals/paper		41.8	37
Naïve vs. Model Animals			
Measuring Development of Sensitization	Effect size (95% CI)	-3.4024 SD (-4.5993 - - 2.2054)	-2.7002 SD (-3.5240 - -1.8764)
	I^2 (%)	83.5	79.2
	No. of experiments	10	18
	No. of animals	208	332
Measuring Expression of Sensitization	Effect size (95% CI)	-1.3757 (-1.9778 - - 0.7736)	-1.3773 SD (-2.0218 - -0.7328)
	I^2 (%)	63.3	71.1
	No. of experiments	14	18
	No. of animals	283	348

7.2.2.6.2.1.2 Results from updating the search

Updating the search by a year added an extra 9 publications. These included 8 experiments describing the development phase of behavioural sensitization in response to chronic cocaine administration and 4 experiments measuring the expression of behavioural sensitization. All experiments that measured the development of behavioural sensitization were model-characterising experiments (8 experiments, 123 animals) and looked at the effect of chronic cocaine administration on the amplification of locomotor activity. A further, 4 experiments (using 65 animals) looked at characterising the model and measured outcome after withdrawal from the chronic drug administration paradigm.

Altogether, model-characterising experiments used 188 animals in total, which were either rats (6 experiments, 101 animals) or mice (6 experiments, 88 animals). All of the rats used were Sprague Dawley rats, while the strain of mice used was much less consistent with 1 experiment using C57 (24 animals), 1 experiment using C57BL/6 (6 animals), 2 experiments using C57BL/6J (19 animals) and another two experiments using C57BL/6J x 129S1/SvImJ mice (39 animals). Most experiments used male animals (7 experiments, 94 animals, 50% of all animals), 2 experiments used female animals (32 animals) and 3 experiments did not state the sex of animals used at all (63 animals). The dose of cocaine administered repeatedly over time to initiate behavioural sensitization ranged from 0.15 mg/kg to 20 mg/kg, administered once a day (10 experiments), every two days (1 experiment) or three times a day (1 experiment). Cocaine was administered in this manner over 5 (2 experiments), 7 (5 experiments), 8 (2 experiments) or 10 days (3 experiments). The time between abstinence from this administration paradigm and measurement of sensitization taken in response to a challenge dose of cocaine was either one, 4, 14 or 51 days.

Once the original dataset was updated with data obtained from the update search, the effect of repeated cocaine administration on locomotion during the development of sensitization gave a pooled effect size of -2.7002 SD units (95% CI -3.5240 - -1.8764), meaning cocaine administered worsened outcome by 2.7 SD units. Substantial heterogeneity was observed in the data ($\tau^2 = 2.0491$, $I^2 = 79.2\%$), but unfortunately, even the updated dataset contained too few data points to be able to perform univariate metaregression. By pooling experiments that measured locomotor activity after withdrawal from a chronic abuse paradigm, a challenge dose

of cocaine had an effect of -1.3773 SD (95% CI -2.0218 - -0.7328) on locomotor activity. This meant that a challenge dose of cocaine worsened behavioural outcome by 1.38 SD units after withdrawal from a chronic drug administration paradigm. This dataset showed a substantial amount of heterogeneity ($\tau^2 = 0.6372$, $I^2 = 71.1\%$) likely seen again due to the small sample size, which also meant univariate meta-regression could not be performed due to low power.

7.2.2.6.2.2 Experiments measuring the effect of a treatment drug

7.2.2.6.2.2.1 Results of the original search

8 experiments (using 142 animals), all from the same publication, measured the effect of a variety of possible therapeutics on the development of sensitization in animals initiated by a chronic cocaine administration paradigm. In comparison, 24 experiments (using 287 animals) had measured the effect of possible therapeutic agents on the expression of behavioural sensitization, tested after withdrawal from a chronic cocaine administration paradigm. Together, all but one of these treatment exploring studies used rats (31 experiments, 409 animals), with 1 experiment using C57BL/6J mice (20 animals). Rats used were mostly Sprague Dawley rats (29 experiments, 391 animals), with 2 experiments using Wistar rats (18 animals). All, but one of the experiments used male animals (409 animals), while the sex of the animals used was not specified in that one experiment (20 animals). Dose of repeated cocaine administrations before withdrawal from the drug ranged from 10mg/kg – 30 mg/kg, with 30 mg/kg being the most prevalent dose given in these experiments (22 experiments). Cocaine was administered once daily either over 5 (3 experiments), 7 (22 experiments), 10 (6 experiments) or 14 days (1 experiment). Measurements of locomotion were taken after administration of a challenge dose of cocaine after withdrawal periods of 5 (3 experiments), 10 (7 experiments), 21 (8 experiments) or 22 days (6 experiments). Therapeutic agents were administered either once (9 experiments), 5 (2 experiments), 7 (7 experiments), 8 (8 experiments) or 10 (6 experiments) times during behavioural sensitization. Treatment was either given for the duration of the repeated administration of cocaine or just before the challenge cocaine dose was administered.

The overall effect seen in studies measuring locomotion during the development of behavioural sensitization was that agents administered were not able to counteract the amplified locomotor activity produced by prolonged cocaine administration.

Overall, there was actually an increase in locomotor activity (i.e. worsening in behavioural outcome) by 1.0149 SD units (95% CI -3.1163 to 5.1460, Table 7.5). This, of course, is strongly affected by the fact that all the data came from a single study and overall amounted only to a small sample of experiments. The dataset also showed a substantial amount of heterogeneity ($\tau^2 = 22.8659$, $I^2 = 94.4\%$), possibly due to the small sample size, which also meant that data were not sufficient to explore using univariate meta-regression. Therapeutic agents showed a small improvement in overall behaviour when locomotion was measured after withdrawal from the drug paradigm, and decrease in cocaine-induced hyperactivity by 0.2087 SD units (95% CI -0.2646 - 0.6820) during the expression phase of behavioural sensitization. The dataset showed a moderate amount of heterogeneity ($\tau^2 = 0.4621$, $I^2 = 52.5\%$), but unfortunately, the sample of experiments was still too small to perform univariate meta-regression on.

Table 7.5 A comparison of the data extracted from studies testing treatment drugs on chronic cocaine-induced animal models in the original search and in the updated search.

Chronic locomotor activity		Original Search	Overall with Updated Search
Total no. of publications		22	31
Animals/paper		41.8	37
Model vs. Treated Animals			
Measuring Development of Sensitization	Effect size (95% CI)	-1.0149 SD (-5.1460 - 3.1163)	-1.0149 SD (-5.1460 - 3.1163)
	I^2 (%)	94.4	94.4
	No. of experiments	8	8
	No. of animals	142	142
Measuring Expression of Sensitization	Effect size (95% CI)	0.2087 (-0.2646 - 0.6820)	0.2662 SD (-0.1573 - 0.6896)
	I^2 (%)	52.5	50.7
	No. of experiments	24	27
	No. of animals	287	323

7.2.2.6.2.2.2 *Results from updating the search*

Updating the search by a year added an extra 9 publications, which contributed only 3 additional experiments looking at the effect of therapeutic agents on repeated administration of cocaine. The 3 experiments, reported in the same publication looked at the effect of fluoxetine, clozapine and haloperidol on the expression of behavioural sensitization in response to a challenge drug of cocaine following withdrawal from a chronic cocaine abuse paradigm. As they were from the same laboratory, they all used male, ICR mice. Experimental setup involved administration of 20mg/kg of cocaine over 5 days to the animal in order to initiate behavioural sensitization. Challenge dose of cocaine was administered and locomotor activity measured 19 days after abstinence from chronic cocaine administration paradigm. Therapeutic agents were administered 3 times, each starting 7 days before and finishing 3 days before challenge cocaine administration.

When the effect of therapeutic agents was measured on this expression of sensitization induced by a challenge dose of cocaine, it was found that hyperlocomotion induced by cocaine could be reduced by 0.2662 SD units (95% CI - 0.1573 - 0.6896) with moderate heterogeneity ($\tau^2 = 0.4287$, $I^2 = 50.7\%$).

7.2.3 Are we capturing all the publications of relevance in our search: the impact of alternate search terms

Next, I wanted to test the impact of search terms used on the results of my meta-analysis for studies answering the same research question. I was interested in analysing the quality of my initial search words used to identify publications in this systematic review, as psychosis is a uniquely human and heterogeneous condition that is diagnosed in the clinic through self-reporting and manifests through a complex group of symptoms. Considering we can't assess self-reporting in animals and it is expected that only a certain group of symptoms are being modelled by an animal instead of the full extent of the disorder (Nestler and Hyman, 2010), it is of interest to see how well we are able to identify relevant publications in the first place. The main aims were to assess if using more specific search terms to the animal experiments of interest instead of the human condition, would pick up on more and potentially different publications and ultimately whether missing these in my original search would impact the results gained.

7.2.3.1 Methods

This work was carried out as part of a BSc Biological Sciences Honours dissertation project and I would like to acknowledge Angus Sinclair (AS) for his work on performing the update search, screening of studies for inclusion and exclusion and extracting data for included publications. Data were checked, and meta-analyses were ran by myself.

7.2.3.2 Search strategy

Publications of relevance were identified from publications included and described in Chapter 3, by filtering using the word “cocaine” in title, abstract and key words of publications. This was pooled with a set of publications identified through an update search of PubMed, completed in January 2015, as described above in section 7.2.2. This pooled dataset will be hereon referred to as the “Original search”, as these publications were identified using the original search string specified in Chapter 2. To look at the effect of an alternate search strategy, an alternative search string was created, which was to identify cocaine-induced locomotor behaviours in animals using more specific search criteria. This alternative search used the following search criteria:

[cocaine] AND [motor act* OR hyperact* OR hyperkinesis OR climbing OR rearing OR behavioural sensitisation OR head shak* OR head twitch* OR prepulse OR prepulse OR acoustic startle OR acoustic reflex OR startle reflex OR auditory startle response OR latent inhibition OR social withdrawal OR motor inhibition OR catalep* OR nesting behaviour OR nest building OR stereotyp* OR sensory gating]

Results were filtered using the same animal filter as mentioned in Chapter 2 (Hooijmans et al., 2010b). The resulting dataset is hereon referred to as the “Alternate search”. All publications were screened independently by title and abstract in MS Access by two reviewers (FC and AS) and discrepancies were resolved by a third screener (myself).

7.2.3.3 Inclusion criteria

The only variable of interest was the effect of using an alternate set of search terms on the data obtained. Therefore, all methods used for inclusion were the same as described above in section 7.2.2 of this chapter. To reiterate, initial methods used

for screening for inclusion and exclusion were identical to those described in Chapter 2, with the addition of the following criteria: Only experiments describing the effects of cocaine on locomotion were included. Studies specifically investigating the addictive effects of cocaine, those testing cocaine in transgenic mice and those that used animals with co-morbidities were also excluded. Experimental paradigms using both acute and chronic administration of cocaine were included, and this was not restricted to any particular route of delivery. The only outcome measure was horizontal locomotion and therefore other outcome measures, such as stereotyped behaviour, were not extracted.

7.2.3.4 Data extraction

Data extraction was also identical to those described in section 7.2.2 of this chapter. From each study included, experimental comparisons describing the effects of acute or repeated administration of cocaine on locomotion and those exploring modulation of this effect through the administration of therapeutic agents were included. These comparisons were extracted and analysed separately. As before, a single reviewer (AS) extracted study design, quality and outcome data for each included comparison as described in Chapter 2. For experiments describing acute effects of cocaine, where data were presented as total locomotion over a period of time, total movement and time period of assessment were extracted. Where locomotion was reported over a period of time through activity at multiple time points, the mean activity over that time was taken and time period of assessment was recorded as final time point of assessment. In experiments describing chronic cocaine administration paradigms, where the development of behavioural sensitization was measured by reporting locomotor activity on first and last day of cocaine administration, the difference between these two measures was taken. For experiments measuring the expression of behavioural sensitization by reporting locomotor activity in response to a challenge of cocaine after withdrawal from the repeated administration paradigm, the first withdrawal time point was taken. 15% of all data extracted was checked by me and any errors that were encountered within this set were changed. As no consistent errors were identified within this subset of the data, everything was taken ahead for further analysis. Overall, both raw data and analyses were also subject to a sense check and any data that appeared erroneous was double-checked within the publication.

7.2.3.5 Analysis

Meta-analysis methods were as described in Chapter 2. Acute cocaine-induced locomotor activity and chronic cocaine-induced sensitized locomotor activity were analysed separately. Moreover, for chronic cocaine administration paradigms, experiments measuring the development of behavioural sensitization and later measurements of expression of behavioural sensitization in response to a challenge dose of cocaine, were also analysed separately. To correct for multiplicity of testing a Holm-Bonferroni adjusted critical p value was calculated to account for the number of variables tested within subgroup analyses, calculated separately for study quality and study design items. For study quality items and for study design items explored in experiments using chronic cocaine administration paradigms adjusted critical p value was set at $p < 0.006$. For study design items explored in experiments using acute cocaine administration paradigms and characterising the model adjusted critical p value was set at $p < 0.009$ and for those looking at the effect of a treatment on models of cocaine-induced psychosis it was set at $p < 0.007$. This included variables of study quality and study design characteristics as described and explained above in section 7.2.2 of this chapter.

7.2.3.6 Results

In total, the alternate search for cocaine-induced locomotor activity identified 2831 publications and screening of these results narrowed this down to 824 relevant publications. Unfortunately, due to time constraints and the large number of publications identified, data extraction was only possible for the most recent publications published up to and including 2010. This gave a final dataset of 85 included publications. For this reason, in order to make the comparison between the data from the original search and the alternate search comparable, data obtained through data extraction from publications in our original search was also limited to the same publication date. This meant that the original search included data from 28 publications, which were published in or after 2010.

Reporting of measures taken to reduce bias and other methodological quality criteria within publications were more prevalent for some items than others (Table 7.6). In the original dataset the median number of study quality checklist items scored was 2 (interquartile range [IQR] 1-2). Random allocation of animals to group was reported in 8 publications (29%), blinded assessment of outcome in 4 (14%), reporting of

animals excluded from analysis in 1 (4%), a statement of potential conflict of interest in 13 (46%), and compliance with animal welfare regulations in 26 (93%). No publications reported having carried out a sample size calculation, the availability of a protocol for their study or having blinded the induction of model or administration of treatment. Overall, the prevalence of reporting of these items assessed was very similar between the two different searches. A substantial difference between the two searches was only seen in the reporting of compliance with animal welfare regulations.

Table 7.6 A comparison of the reporting of methodological quality items and measures taken to reduce risks of bias in studies identified in our original search, in the alternate search and with all the data pooled.

	Original Search	Alternate Search	Overall
Total No. of publications identified	28	85	113
Number of publications reporting ... (Overall %)			
Random allocation to group	8 (29%)	22 (26%)	30 (27%)
Allocation concealment	0	0	0
Blinded assessment of outcome	4 (14%)	11 (13%)	15 (13%)
Sample size calculation	0	0	0
Compliance with Animal Welfare Regulations	26 (93%)	85 (100%)	111 (98%)
Statement of potential conflict of interest	13 (46%)	39 (46%)	52 (46%)
Prespecified or explanation of exclusion of animals	1 (4%)	2 (2%)	3 (3%)
Availability of a study protocol	0	0	0
Median quality (/8)(interquartile range)	2 (1-2)	2 (1-2)	2 (1-2)

7.2.3.6.1 Acute cocaine-induced locomotor activity

7.2.3.6.1.1 Results from the original search

In the original dataset 34 experiments reporting data for the effects of cocaine on locomotion (using 591 animals) were identified and extracted. 21 of these experiments were performed in rats (332 animals) and 13 using mice (259 animals).

The most commonly used strain was Sprague Dawley (18 experiments, 302 animals, representing 51% of all animals used and 91% of rats). There was much less consistency in the strain of mouse used, the most common being C57Bl/6 mice (5 experiments, 99 animals, representing 38% of mice). Most experiments used male animals (27 experiments, 438 animals), while one (16 animals) used female animals, and 6 (140 animals) did not state the sex of animal used. The dose of cocaine used in the experiments varied from 0.15 mg/kg to 30 mg/kg, with 26 experiments (76%) using 15mg/kg or less.

Altogether, administration of cocaine increased locomotor activity by 0.8165 SD units (95% CI 0.5003 to 1.1326, Table 7.7). As before, this is defined as a worsening in behavioural outcome, and therefore an increased efficacy of the model at showing signs of hyperactivity thought to model the psychotic symptoms seen in humans.

Table 7.7 A comparison of the data extracted from studies reporting acute cocaine-induced locomotor activity in the original search and in the alternate search. Negative values indicate a worsening in outcome and positive values indicate an improvement in the outcome.

Acute locomotor activity		Original Search	Alternate Search
No. of publications		17	71
Animals/paper		41	41
Naïve vs. Model Animals	Effect size (95% CI)	-0.8165 SD (-1.1326--0.5003)	-1.5891 SD (-1.7643--1.4138)
	<i>I</i> ² (%)	53.1	48.2
	No. of experiments	34	99
	No. of animals	591	1674
Model vs. Treated Animals	Effect size (95% CI)	0.8171 SD (-0.2139-1.8482)	0.4027 SD (0.1581-0.6474)
	<i>I</i> ² (%)	65.4	61.4
	No. of experiments	7	86
	No. of animals	100	1207

The dataset also showed moderate heterogeneity ($\tau^2 = 0.3867$, $I^2 = 53.1\%$), which was only significantly explained by the dose of cocaine used to induce the model ($p = 0.0046$, $\tau^2 = 0.1864$, $I^2 = 35.92\%$, $\text{adj } R^2 = 51.79\%$, Figure 7.10). Heterogeneity was not explained by any of the other study characteristic variables investigated with univariate meta-regression, namely: animal, strain and sex of animals, dose of cocaine used to induce the model, method of cocaine administration and time of outcome assessment.

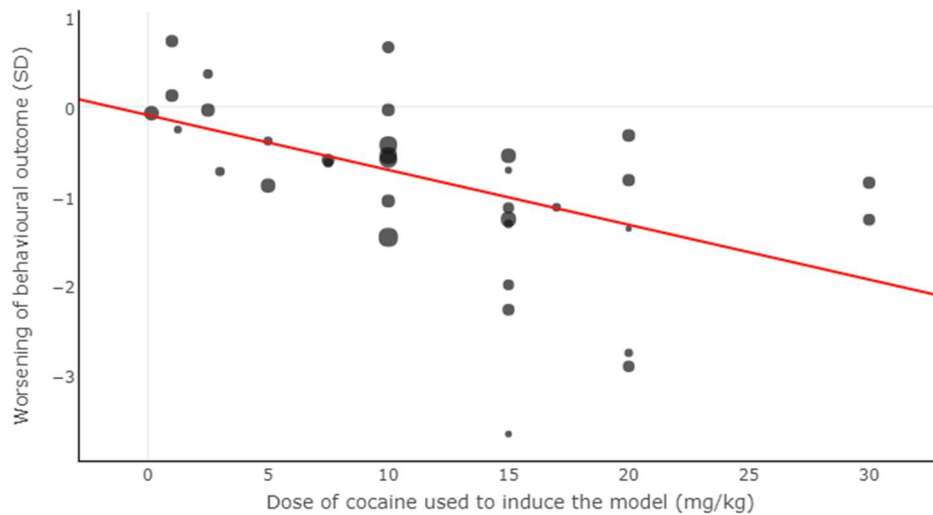


Figure 7.10 Relationship between acute dose of cocaine used to induce the model and reported effect size in model-characterising experiments of the

Limiting studies according to publication date limit, unfortunately, excluded many previously included data from experiments testing the efficacy of interventions in moderating cocaine-induced hyperlocomotion. The comparison set included seven experiments from the original search dataset (using 100 animals). All of these experiments were performed in male rats, using Long-Evans (1 experiment, 15 animals), Sprague Dawley (5 experiments, 73 animals) or Wistar (1 experiment, 12 animals) rats. In four experiments animals were given 10mg/kg dose of cocaine to induce model, in two 15mg/kg, and in one experiment animals were administered 20mg/kg to induce the model. Time of behaviour assessment ranged from 30 minutes to 2 hours. Time of treatment administration was as early as a month before cocaine administration in some cases, but otherwise, it was usually administered alongside cocaine. Drugs tested as treatments included the typical antipsychotic flupenthixol, antidepressants, dopamine D3R agonists and an inhibitor of protein palmitoylation.

Considered together, drug treatments improved hyperlocomotion induced by cocaine by 0.8171 SD units (95% CI -0.2139-1.8482), with again moderate heterogeneity ($\tau^2 = 0.7989$, $I^2 = 65.4\%$). Due to the extremely small sample size, univariate meta-regression could not be performed.

7.2.3.6.1.2 Results from the alternate search

Data extraction from publications identified using the alternative search strategy established 71 publications exploring the effects of a single administration of cocaine on locomotor activity. From these publications, 99 experiments were established to be testing the ability of cocaine in increasing locomotor activity using 1674 animals.

52 of these experiments used mice as their animal model of choice (819 animals), 45 experiments used rats (835 animals) and two used marmosets (20 animals). The most common strain used for mice was CD-1 (12 experiments, 170 animals, 21% of all strains of mice and 10% of all animals used), while the most common strain for rats was Sprague Dawley (25 experiments, 464 animals, 56% of rats used and 28% of all animals used). The majority of animals used, here as seen in the other datasets, was male (88 experiments, 1483 animals), while only 6 experiments used female animals (104 animals) and 4 experiments used both (77 animals). The dose of cocaine used to induce the model varied from 1.25 mg/kg to 66 mg/kg. Duration of behavioural assessment after cocaine administration varied from 5 minutes to 4 hours and 20 minutes.

Overall, these model-characterising experiments showed an overall increase in locomotion by 1.5891 SD units (95% CI 1.4138-1.7643) in animals in response to cocaine, with moderate heterogeneity ($\tau^2 = 0.3590$, $I^2 = 48.2\%$). Variables that significantly accounted for a proportion of the heterogeneity included strain of animal used within experiments ($p = 0.0034$, $\tau^2 = 0.2263$, $I^2 = 35.81\%$, $\text{adj } R^2 = 36.96\%$, Figure 1) and dose of cocaine used to induce the model ($p = 0.0007$, $\tau^2 = 0.2939$, $I^2 = 44.18\%$, $\text{adj } R^2 = 18.15\%$, Figure 7.11). No other variables of internal validity or study characteristics significantly accounted for heterogeneity seen in the dataset.

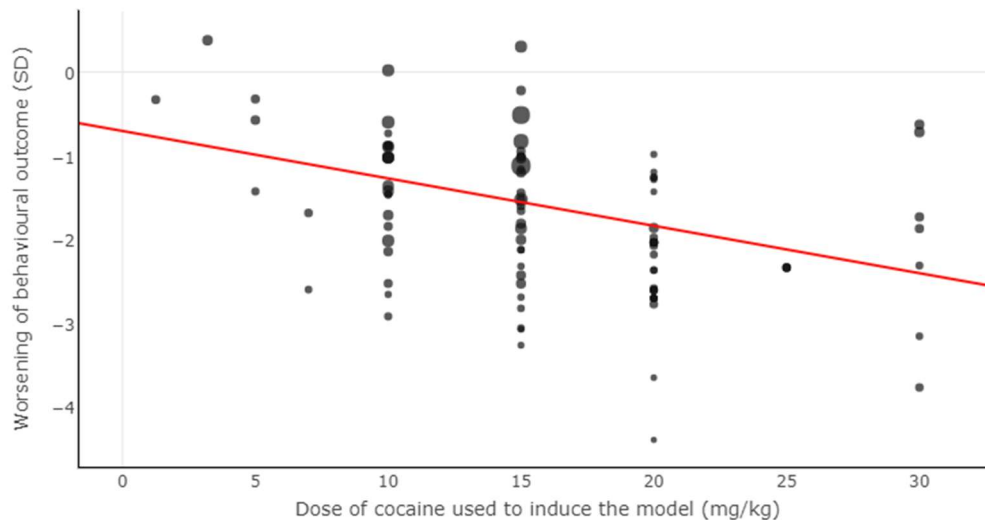


Figure 7.11 Relationship between acute dose of cocaine used to induce the model and reported effect size in model-characterising studies from the alternate search.

86 experiments also tested the efficacy of interventions in moderating cocaine-induced hyperlocomotion (1207 animals). The two most commonly used animals in these experiments were mice (45 experiments, 641 animals), and rats (40 experiments, 556 animals), in addition to one experiment using marmosets (10 animals). Experiments using mice used 8 different species to create the models, but most common models were Swiss-Webster (13 experiments, 137 animals, 21% of all mice used and 11% of all animals used) and CD-1 (10 experiments, 137 animals, again, 21% of all mice used and 11% of all animals used). For rats, the most common strain used was Sprague Dawley (22 experiments, 370 animals, 67% of all rats used and 31% of all animals used). In terms of sex used, the same pattern was seen as before. The majority of experiments used male animals (83 experiments, 1161 animals, 96% of all animals), while one used both (10 animals) and 2 did not specify the sex of the animals at all (36 animals). The dose of cocaine used to induce the model varied from 5 mg/kg to 66 mg/kg, with the majority of experiments administering 20 mg/kg or lower doses to animals (79% of all experiments). Duration of behavioural assessment after cocaine administration varied from 10 minutes to 4 hours and 20 minutes. Time of treatment administration varied from over two months before model induction, to administration at the same time as cocaine. Overall treatment drug studies improved cocaine-induced hyperactivity by 0.4027 SD units (95% CI 0.1581-0.6474) with slightly more observed heterogeneity than seen in the data identified in the original search, however, still only moderate ($\tau^2 = 0.5700$, $I^2 = 61.4\%$). None of the variables assessed in univariate meta-regression significantly accounted for any of this heterogeneity.

7.2.3.6.2 Chronic cocaine administration schedule

7.2.3.6.2.1 Experiments characterising the model

7.2.3.6.2.1.1 Results of the original search

When limiting the original search results on cocaine studies to publications published in or after 2010, 32 experiments using 219 animals reported measurements of locomotor activity after chronic cocaine administration. 12 experiments measured the development of behavioural sensitization through repeated cocaine administration (219 animals). An additional 11 experiments (using 245 animals) reported the effect of a challenge dose of cocaine on locomotion in animals who had previously been sensitized to the drug through a chronic drug administration paradigm.

In total these experiments used 464 animals, most of which were mice (15 experiments, 324 animals), while 8 experiments reported outcomes in rats (141 animals). The most commonly used strain of mouse was C57BL/6J (6 experiments, 111 animals, 34% of all mouse strains used, 24% of all animals used). All rats in the dataset were Sprague Dawley rats. The majority of these animals were male (15 experiments, 296 animals), with two experiments using female animals (32 animals), one experiment using both sexes (22 animals), and 5 experiments not stating the sex of animals used at all (115 animals). Dose of cocaine administered repeatedly ranged from 0.15 mg/kg to 20 mg/kg, with most experiments using a chronic dosage paradigm of 10 mg/kg (7 experiments) or 15 mg/kg (12 experiments) doses given repeatedly. These doses were either given every two days (1 experiment), once a day (21 experiments) or three times a day (1 experiment) over 5 (9 experiments), 7 (5 experiments), 8 (6 experiments), or 10 days (3 experiments). Expression of behavioural sensitization measured in the 11 experiments by assessing locomotor activity in response to a challenge dose of cocaine, was measured after a withdrawal period ranging between 1-51 days after the last day of the chronic drug paradigm.

Overall these studies indicate that chronic cocaine administration amplifies the hyperlocomotor response to cocaine by 2.2028 SD units (95% CI 1.2898-3.1157, Table 7.8), implying the development of behavioural sensitization. Substantial heterogeneity was observed in the data ($\tau^2 = 1.4818$, $I^2 = 76.4\%$). Administration

of a challenge dose of cocaine after previous chronic exposure to the drug increased locomotor activity by 1.3654 SD units (95% CI 0.6969 – 2.0338) compared to healthy, control animals when taking these experiments together. Substantial heterogeneity was observed in the data ($\tau^2 = 0.6918$, $I^2 = 70.2\%$), however, neither of the datasets was sufficiently big enough to perform univariate meta-regression.

Table 7.8 A comparison of the data extracted from studies reporting chronic cocaine-induced locomotor activity in the original search and in the alternate search,

Chronic locomotor activity		Original Search	Alternate Search	Pooled dataset
Total no. of publications		14	29	43
Animals/paper		33	24	26
Naïve vs. Model Animals				
Measuring Development of Sensitization	Effect size (95% CI)	-2.2028 SD (-3.1157--1.2898)	-2.1571 SD (-3.0900--1.2243)	-2.1619 SD (-2.8149--1.5090)
	I^2 (%)	76.4	82.3	80.2
	No. of experiments	12	24	36
	No. of animals	219	351	534
Measuring Expression of Sensitization	Effect size (95% CI)	-1.3654 SD (-2.0338--0.6969)	-1.465 SD (-1.9836--0.9465)	-1.4231 SD (-1.8089--1.0373)
	I^2 (%)	70.2	60.9	63.5
	No. of experiments	11	22	33
	No. of animals	245	340	585

7.2.3.6.2.1.2 Results from the alternate search

Performing the search using a set of alternate search words identified an additional 84 experiments reporting experimental paradigms using chronic cocaine administration to measure both the development and expression of behavioural sensitization in response to the drug. 24 of these experiments (using 351 animals) measured cocaine-induced locomotor activity at the end of the development phase

of behavioural sensitization initiated by chronic cocaine administration, and 22 experiments (using 340 animals) measured locomotion after withdrawal from this administration paradigm in response to a challenge dose of cocaine.

Overall, animals used (658 in total) were mainly mice (19 experiments, 306 animals) and rats (24 experiments, 357 animals), but also three experiments reported using marmosets (28 animals). The most common strain of mouse used was C57BL/6J mice (7 experiments, 115 animals, 38% of all mice used), and for rats Sprague Dawley was the most prevalent strain of choice in this dataset (18 experiments, 263 animals, 74% of all rats used). Most of the animals used were male (37 experiments, 566 animals, 82% of all animals used), with 4 experiments reporting locomotion in female animals (54 animals) and 3 reporting using animals of both sex (51 animals), with the remaining two experiments not specifying sex of animals used (20 animals). Development of behavioural sensitization was initiated using doses of cocaine ranging from 5 mg/kg to 30 mg/kg, repeatedly administered either once, three times a day, or every other day. Duration of chronic cocaine administration ranged from 3 – 21 days between different experiments. Most experiments administered 10 mg/kg (11 experiments) or 15 mg/kg (24 experiments) of cocaine using a once a day paradigm (36 experiments, 78% of all experiments) for 5 (13 experiments) or 7 days (16 experiments). For experiments measuring withdrawal from this chronic cocaine paradigm, challenge dose and subsequent measurement of locomotion was performed 1 to 42 days after last cocaine administration.

Overall experiments measuring locomotor activity at first and last day of chronic cocaine administration schedule showed that cocaine administered after behavioural sensitization had been initiated locomotor activity was increased by 2.1571 SD units (95% CI 1.2243 – 3.0900). Heterogeneity was quantified to be substantial ($\tau^2 = 3.3068$, $I^2 = 82.3\%$). Experiments that measured locomotor activity after a period of withdrawal show that locomotor activity increased in response to a challenge dose of cocaine by 1.465 SD units (95% CI 0.9465 – 1.9836) compared to animals that had not been previously administered cocaine. Heterogeneity was only moderate for this subset of studies ($\tau^2 = 0.6377$, $I^2 = 60.9\%$). Due to low sample of experiments contributing to both meta-analyses, heterogeneity could not be explored further in either datasets using meta-regression.

7.2.3.6.2.2 *Experiments measuring the effect of a treatment drug*

7.2.3.6.2.2.1 *Results of the original search*

No experiment measured the ability of therapeutic drugs to modulate the development of behavioural sensitization. 9 experiments looked at the reversal of the expression of behavioural sensitization through the administration of therapeutic drugs and their ability to modulate exaggerated hyperactivity in response to a challenge dose of cocaine following a previous chronic administration paradigm of the drug. These experiments used again mice (3 experiments, 36 animals, ICR strain) or rats (6 experiments, 52 animals, Sprague Dawley strain) in their experimental designs. All animals reported were male. Cocaine administered repeatedly to induce development of behavioural sensitization was either given at a dose of 20 mg/kg (3 experiments) or 30 mg/kg (6 experiments), once a day over 5 (3 experiments) or 7 days (6 experiments). Expression of this behavioural sensitization and ability of therapeutic agents to reverse this was measured after withdrawal for 19 or 22 days from this paradigm. Therapeutic drugs were administered either once 2 hours before challenge dose of cocaine was given or three times starting 7 days before to 20 hours before challenge cocaine administration.

These agents overall were able to reduce the exaggerated hyperactivity induced by a challenge dose of cocaine after previous chronic administration to the drug by 1.2658 SD units (95% CI 0.5041 - 2.0276, Table 7.9) with very little heterogeneity ($\tau^2 = <0.0001$, $I^2 = 21.9\%$).

Table 7.9 A comparison of the data extracted from studies reporting treatment effects on chronic cocaine-induced locomotor activity in the original search and in the alternate search.

Chronic locomotor activity		Original Search	Alternate Search	Pooled dataset
Total no. of publications		2	21	23
Animals/paper		44	26	27
Treatment vs. Model Animals				
Measuring Development of Sensitization	Effect size	-	1.0198 SD	1.0198 SD
	(95% CI)	-	(0.0090 - 2.0306)	(0.0090 - 2.0306)
	I^2 (%)	-	87.3	87.3
	No. of experiments	-	24	24
	No. of animals	-	327	327
Measuring Expression of Sensitization	Effect size	1.2658 SD	0.9569 SD	1.0937 SD
	(95% CI)	(0.5041 - 2.0276)	(0.2609 - 1.6529)	(0.5744 - 1.6131)
	I^2 (%)	21.9	71	63.9
	No. of experiments	9	15	24
	No. of animals	88	217	305

7.2.3.6.2.2.2 Results from the alternate search

Through the alternate search a further 39 experiments (using 544 animals) were identified to be exploring the effect of potential therapeutic agents on chronic administration paradigms of cocaine. 24 experiments looked at the effect of these compounds on the development of behavioural sensitization (327 animals) or tried to reverse its expression (15 experiments, 217 animals). These experiments similarly to model-characterising studies used either rats (20 experiments, 290 animals), mice (18 experiments, 246 animals) or marmosets (1 experiment, 8 animals). The most common rat strain used was Sprague Dawley rats (12 experiments, 162 animals, 56% of all rats, 30% of all animals used) and the most common strains of mouse used were C57BL/6 and C57BL/6J (5 experiments, 82

animals, 33% of all mice, each). Marmosets were all common marmosets. Sex of animals used was overwhelmingly male (34 experiments, 464 animals, 85% of all animals), with one experiment reporting the use of both male and female animals in their experimental groups (8 animals), and 4 experiments not stating the sex of animals used (72 animals). Dose of cocaine repeatedly administered to create the model ranged from 5 mg/kg to 20 mg/kg, with most experiments administering 10 mg/kg (13 experiments) or 15 mg/kg (19 experiments). These doses were administered mostly once a day (29 experiments), but also some experimental paradigms administered cocaine repeatedly three times a day, every 2 or every 3 days, over a period of 3 – 14 days. Most commonly cocaine was administered over a course of either 5 days (10 experiments) or 7 days (14 experiments). For experiments that measured locomotor activity after withdrawal from this chronic administration paradigm, challenge dose of cocaine was given and measurement was taken after abstinence from the drug ranging from 1 – 20 days.

Therapeutic agents were able to reduce the increase in locomotor activity seen in animals as a result of chronic administration of cocaine during the development phase of behavioural sensitization by 1.0198 SD units (95% CI 0.0090 – 2.0306), with substantial heterogeneity observed in the data ($\tau^2 = 4.3302$, $I^2 = 87.3\%$). Treatment compounds tested dampened the hyperactivity induced by administration of a challenge dose of cocaine after withdrawal from the chronic cocaine administration paradigm by 0.9569 SD units (95% CI 0.2609 - 1.6529), with again, substantial heterogeneity observed in the data ($\tau^2 = 1.0599$, $I^2 = 71.0\%$). Unfortunately, like before, the low sample of experiments contributing to both meta-analyses meant that heterogeneity could not be explored further in either datasets using meta-regression.

7.2.3.6.2.3 Results from pooling all the data

Overall effect sizes for data stratified by search criteria (original search or alternate search) were similar. By pooling data from both searches, chronic cocaine administration showed an overall increase in locomotor activity after the development of sensitization by 2.1619 SD units (95% CI 1.5090 – 2.8149) which was not considerably different from the effect seen by only including data identified through the original search in the meta-analysis (2.2028 SD units with 95% CI 1.2898-3.1157). Pooling the two datasets however, did increase the amount of statistical heterogeneity observed in the overall dataset ($\tau^2 = 2.4790$, $I^2 = 80.2\%$),

when compared to the original dataset ($\tau^2 = 1.4818$, $I^2 = 76.4\%$). While the original dataset was insufficient to perform meta-regression using all variables of interest, I was able to use univariate meta-regression on the pooled dataset. Strain of animal used was the only variable to contribute significantly to heterogeneity ($p = 0.001$, $\tau^2 = 0.7918$, $I^2 = 58.96\%$, $\text{adj } R^2 = 68.06\%$,

Figure 7.12).

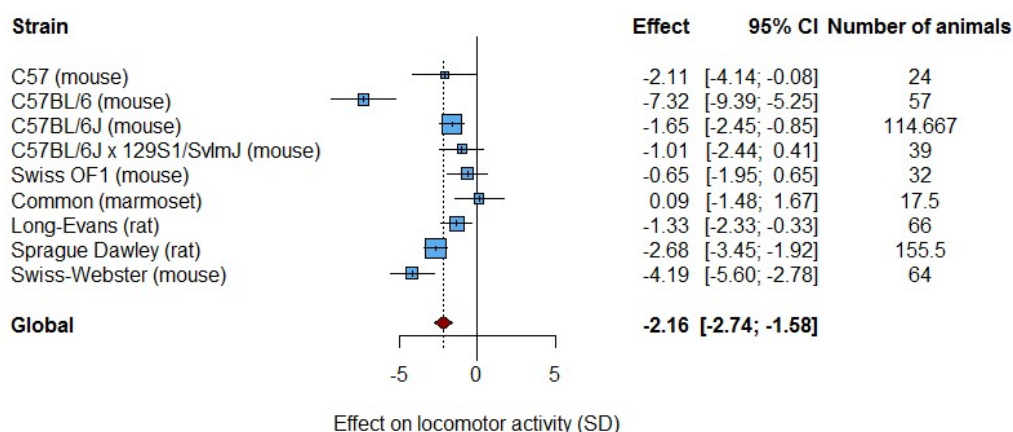


Figure 7.12 Relationship between the strain of animal used and reported effect size in chronic model-characterising studies from both original and alternate searches

Pooling data for experiments that had measured locomotor activity after withdrawal from the chronic cocaine administration paradigm, showed that a challenge dose of cocaine worsened behavioural response by 1.4231 SD units (95% CI 1.0373 - 1.8089). This was a small increase in the effectiveness of the model from that seen in the meta-analysis of only data from the original search (worsening in behaviour by 1.3654 SD with 95% CI 0.6969 - 2.0338). Pooling the two datasets decreased the amount of statistical heterogeneity observed in the overall dataset ($\tau^2 = 0.6139$, $I^2 = 63.5\%$), when compared to the original dataset ($\tau^2 = 0.6918$, $I^2 = 70.2\%$). Strain of animal used within each experiment did not make a significant contribution to heterogeneity in this dataset.

Other variables that did not contribute significantly to heterogeneity in either of the analyses included: reporting of random allocation of animals to groups, allocation concealment, blinded assessment of outcome, specified inclusion or exclusion of animals, a conflict of interest statement, compliance with animal welfare regulations, availability of a study protocol and sample size calculation, species, strain or sex of

animals used, dose of cocaine repeatedly administered, total number of administrations, day of last administration and number of days of abstinence from cocaine for experiments which measured locomotor activity after withdrawal.

No data were identified through the original search describing the effects of treatment compounds on the development of sensitization. Exploring the administration of therapeutic agents on the expression of behavioural sensitization, measured after withdrawal from the model showed an improvement in behavioural outcome by 1.0937 SD units (95% CI 0.5744 - 1.6131). This was a decrease from the effect seen just using data from the original search (1.2658 SD with 95% CI 0.5041 - 2.0276). The amount of statistical heterogeneity observed was more ($\tau^2 = 0.8149$, $I^2 = 63.9\%$) than found in the original search dataset ($\tau^2 = <0.0001$, $I^2 = 21.9\%$). None of these datasets were sufficient to perform univariate meta-regression and therefore heterogeneity could not be explored further.

7.3 Discussion

In this chapter, I have shown that capturing all the data relevant to a specific research question can be difficult. While reporting bias is not as substantial in the subset of the field reviewed as expected, identification of additional studies through updating a search and performing an alternate search show that some relevant publications might have been missed through my systematic review.

7.3.1 Publication bias and selective reporting bias

Data collected in the context of experiments reporting developmentally induced animal models of schizophrenia failed to show substantial evidence for possible publication bias in the field for both model-characterising and treatment-testing experiments. On analysis of the entire dataset of model-characterising experiments calculations suggested that a potential 458 experiments were missing from the literature and that the overall reported effect of model induction on behaviour was being underestimated. Removing outliers in data during sensitivity analysis, still suggested some extent of publication bias using Egger's regression, however, no asymmetry was detected using trim-and-fill. These disparities in outcome before and after sensitivity analyses suggest that extreme values in the dataset associated with small sample sizes used within these experiments, show why sometimes animal studies can be unreliable and imprecise when based on groups of small sizes. It is

suggested that when the size of a sample is small the variance of this sample will be further from the underlying population variance, which introduces a measurement error and thus may weaken the SMD approach to calculation of effect sizes. In fact a recent study shows that the statistical power of SMD for calculating effect sizes for animal experiments using small sample sizes is not as high as it is for calculating this using the NMD approach (Wang et al., 2018). Perhaps a review of the entire field might yield different results in future, but it appears that in the context of the dataset reviewed here there is no substantial evidence for the presence of publication bias.

This is very interesting of course as other pre-clinical fields show that publication bias is a substantial issue (Sena et al., 2010). However, a recent analysis of publication bias assessment in animal studies shows that publication bias may have been overestimated in some previous preclinical studies. This is due to their use of standard error-based precision estimates in assessment of publication bias instead of the sample size-based precision estimates used here, which recent research suggests causes distortion of funnel plots and can lead to false-positive results (Zwetsloot et al., 2017). This distortion has been found to be largest for experiments of small sample sizes and therefore relevant for all assessments looking at preclinical studies. If there is in fact no publication bias in the psychosis research field, perhaps this can be explained by the fact that our understanding of the disorder is still very limited and the need for new treatment drugs is great. This might mean that more studies are published in this field of research compared to others, regardless of overall directionality of outcomes.

Assessment of selective reporting bias on the other hand in the field shows that this problem seems to be extensive. The method used here to estimate the extent of this problem is arguably a crude way of estimating the extent of selective outcome reporting bias. Therefore, it is likely that this is only an underestimation of the true extent of this issue as we cannot tell how many outcomes were changed, omitted or introduced at later stages of experiments.

Estimations in clinical trials imply that about 40-62% of all data collected in clinical trials is at risk of this bias (Dwan et al., 2008). The exact extent of this in other pre-clinical research fields is unknown, but evidence shows that publications reporting neurological diseases show an excess of significant studies, likely explained by publication bias and other reporting biases in these research fields (Tsilidis et al.,

2013). In order to get the entirety of a biological picture that is robust and true, we must be able to see and have all the pieces of knowledge available to interpret. An improvement in initiatives such as pre-registering of studies to overcome publication bias and selective outcome reporting play an integral role in this.

7.3.2 Out-of-date search

Updating the current review, a year after the initial search was performed identified an additional 12 publications describing animals given cocaine to induce hyperactivity. The overall trend seen in the reporting of measures taken to reduce the risk of bias and other methodological quality criteria showed little change upon the update of the dataset. The only differences seen were in the prevalence of the reporting of compliance with animal welfare regulations (100% in the update search, compared to 66% in the original search, raising prevalence to 71% in pooled dataset), and in the reporting of any potential conflicts of interest (15% in the original search, 50% in the update search, raising prevalence to a total of 21% in the overall dataset). Interestingly, the reporting of other measures such as blinded assessment of outcome and descriptions of any animal exclusions, were not reported in publications of the update search at all, despite having been described as two of four core items of reporting standards (Landis et al., 2012). This is somewhat surprising considering that the impact of these items in pre-clinical research has gained a lot of attention in recent years (Macleod et al., 2015) and that the ARRIVE guidelines were published in 2010 (Kilkenny et al., 2010b), almost a whole 4 years before any of the publications in the update search. On the other hand, it has been shown by other research that the impact of these guidelines appears to be slow and even an assessment of its impact in the year of which the original search was performed showed that it was not well implemented (Baker et al., 2014).

Improvements in light of these guidelines seems to advance slowly in other fields also (Bahor et al., 2017; Gulin et al., 2015). Implementing changes in editorial policy in order to increase the completeness of reporting of study design has potential to show greater improvements in future reporting of these measures (Macleod, 2017).

The number of animals used per publication did not seem to change a great deal between the original and the updated search. Overall the number of experiments using acute administration paradigms was much more common than those using repeated exposure paradigms in the original search (157 acute cocaine experiments and 56 chronic experiments). In the update search, the number of publications

reporting experiments using acute cocaine administration and chronic cocaine administration to model psychosis were similar (17 acute experiments and 15 chronic cocaine experiments). There seems to be somewhat of a disagreement in the field as to which dosage regimen in animals is likely to have most validity for the clinical condition of psychosis. In humans, it is widely believed that psychotic symptoms are related to chronic consumption of cocaine (Roncero et al., 2013), where a 'binge' often involves short periods of heavy use with periods of little or no use (Myers et al., 1995). The chronic administration paradigm induces a persistent state of sensitization, which is thought to be a better model of the clinical condition (Featherstone et al., 2007). Especially as schizophrenia has been described as a state of endogenous sensitization and this is only induced after a repeated experimental paradigm (Weidenauer et al., 2017). Nevertheless, a review of the clinical literature shows that even acute exposure to psychostimulants is able to elicit psychotic symptoms thus arguing that acute administration models have more predictive validity (Segal et al., 1981). Perhaps the validity of the dosage regimen lies in the actual dosage of cocaine administered. Early research shows that large doses of a single administration of psychostimulants can produce acute psychotic symptoms (Bell, 1973). When used at lower doses, psychotic symptoms only appear in a subset of abusers and do so after repeated use of these drugs (Bramness et al., 2012). In experiments from both the original and update search reviewed here, the dose of cocaine administered did not differ substantially between acute or chronic administration paradigms. In the original search, the dose of cocaine administered acutely ranged from 0.04 mg/kg to 65 mg/kg for all experiments extracted with 88% of experiments using doses of less than 20 mg/kg in model-characterising studies and 99% of experiments using doses of less than 30 mg/kg for treatment exploring experiments. This did not substantially change in experiments extracted as part of the update search, where acute dose of cocaine ranged from 0.15mg/kg to 20 mg/kg. In experiments administering cocaine as part of a repeated experimental setup, doses of cocaine administered fell between similar ranges. Doses ranged from 7.5 mg/kg to 30 mg/kg in all experiments included in the original search, and 0.15 mg/kg to 20 mg/kg in experiments extracted as part of the update search. This seems to model the human pattern of abuse where it has been shown that individuals who are sensitized to the effects of cocaine after chronic use, showed longer regular use of cocaine and less dose escalation over time (Bartlett et al., 1997). Dose did not seem to explain a substantial proportion of the heterogeneity

seen in studies where cocaine was administered chronically, whereas there was a significant relationship between dose of cocaine and outcome measure in acute dosage paradigms. The duration of cocaine administration also differed widely between experiments, but overall, experiments identified through the original search and those identified through the update search showed similar patterns of this experimental setup (repeated cocaine administration ranged from 5-20 days of administration in original search and 5-10 days in update search). In the clinic, the development of psychotic symptoms in individuals who abuse cocaine is positively correlated with the amount and duration of use (Brady et al., 1991). Any significance of differences in duration of cocaine administration or length of withdrawal from drug could not be established on reported outcome in the current analysis.

In terms of the experimental subjects used, this did not change with time. For example, strain of animals used seems to influence effects in behaviour in response to cocaine. Experiments included here from both original and update searches, used Sprague Dawley rats consistently, whereas mouse strains used were more varied. It is clearly established that genetic factors have a strong impact on differences in locomotion seen between different strains in response to both acute and chronic psychoactive drug administration (Ruth et al., 1988; Zombeck et al., 2010). And in fact pre-clinical studies show that locomotor activity in response to the same experimental setup can be very different across different strains (Thomsen and Caine, 2011), showing that central nervous system sensitivity to cocaine is genetically determined (Ruth et al., 1988). This means that in order to get robust results, future studies should validate effects of treatments in multiple strains. Further to this, the sex of animals used in studies identified both in the original and update searches were overwhelmingly just male animals. In fact, all of the experiments looking at the effects of therapeutic agents on both acute and chronic cocaine induced behaviours used male animals. Clinical data show that males are significantly more likely to develop psychosis than females in response to cocaine (Brady et al., 1991), and dopamine release has been shown to be greater after a single administration of a psychostimulant in males as opposed to females (Munro et al., 2006). However, studies also show that the behavioural effects of sensitization as a result of chronic psychostimulant administration are stronger in females (Becker et al., 2001). Nevertheless it is interesting to note that this trend in the use of animals used has not changed in more recent experiments despite more focus on these issues in preclinical research (Check Hayden, 2010).

Overall, the pooled effect of both model-characterising and treatment exploring experiments changed very little by updating the search. This could be due to the small sample size of publications that the update search added to the overall dataset, therefore having little impact on the overall outcome. There was also little difference in observed heterogeneity of the different datasets reviewed. The reporting or not reporting of randomisation was the only variable of interest that seemed to explain part of the observed heterogeneity in acute cocaine experiments before, but not after the addition of the update search.

In clinical systematic reviews, quantitative signals such as changes in statistical significance or relative change in effect magnitude of at least 50% has been used to warrant the need for updating a systematic review (Shojania et al., 2007). In this chapter, we see that by missing data published in the literature a year after the original search has little effect on the overall effect observable. Nevertheless, it can also be argued that additional studies published in subsequent years might add further to not only our overall knowledge, but perhaps also to our confidence in findings already in the literature. For example, knowing that something works or doesn't work in a variety of different experimental paradigms using different species, strains, and sex of animals, or doses of drug administered at various times of treatment administration, is likely to increase the chance of successful translation of information into clinical understanding and practice (Hånell and Marklund, 2014). As the data show there are a large number of drugs, which have been tested in this one model of stimulant-induced psychosis and the relative paucity of evidence for individual drugs suggests that a more systematic approach to drug development may be helpful in future.

7.3.3 Using alternative search methods

Performing the search using alternate search words, more specific to the research question of cocaine's effects on locomotor activity in animals, identified an additional 85 publications. Compared to the 28 publications identified within the same timeframe through the original search, this was a considerable corpus of publications that were missed as part of the original search. The original study used broad search terms related to psychosis, schizophrenia and other psychotic disorders, and it was not specific to induction method nor outcome measure, unlike the alternate search. Therefore, it is likely that additional studies were missed because they were not labelled as being of relevance to psychosis. In fact, many of

these publications claimed to be exploring the addictive properties of cocaine in these animals. As mentioned in Chapter 3, the validity of locomotor activity as a measure of the clinical state of psychosis is debatable. In the clinic humans are not normally hyperactive, however, can often display stereotyped behaviours (Geyer and Moghaddam, 2002).

Moreover, locomotion is a non-specific behaviour in animals and can be interpreted to in many different ways (Pratt et al., 2012). In the context of addiction, acute drug-induced locomotor activity is a measure of an animal's sensitivity to a substance and chronic drug-induced locomotor activity is a measure of incentive-sensitization model of drug craving (Eisener-Dorman et al., 2011). However, locomotor activity in response to psychotomimetic drugs such as cocaine is also widely considered to be a marker of stereotyped behaviours in the clinic such as psychotic agitation (Powell and Miyakawa, 2006), and is used as a robust and consistent measure of the positive symptoms of psychosis (Marcotte et al., 2001)

Despite differences in how studies were labelled and the context they were described in, overall study designs did not differ substantially. Both sets of data used a mixture of mice and rats, with the alternate search also identifying some experiments describing the effects of cocaine in marmosets. The strains of these animals used were also fairly uniform across experiments in the two datasets, whereby most rats used were mainly Sprague Dawley rats, whereas mice were tested in a larger variety of strains. The use of male animals also predominated in both datasets, with female animals being used only in a handful of experiments. To find this pattern among studies that are supposedly modelling addiction in animals is interesting as clinical studies show that there are some striking differences between the sexes in drug abuse (Becker and Hu, 2008). Nevertheless, this shows that the misbalance in reporting of animals of both sexes is an issue in all areas of preclinical research.

Overall the alternate search identified a large number of additional experiments looking at the effects of acute cocaine administration (99 model-characterising and 86 treatment exploring experiments identified in the alternate search compared to only 34 model-characterising and 7 treatment exploring experiments in the original search). Chronic experimental paradigms were less common and described in about equal amount of experiments in the alternate search as the original search. As mentioned already, dependence on a drug tends to develop after repeated, heavier

use, implying that these studies not identified in the initial search do not seem to show any more validity for the clinical abuse patterns seen in the clinic leading to psychosis. Dosage administered was also comparable between the original dataset and the alternate dataset. Within both datasets, experiments characterising the model were more frequently reported than experiments where therapeutic drugs were administered.

In the original dataset, the effect of chronic cocaine administration was more often measured during the withdrawal phase, compared to the alternate dataset where more studies measured the effect of chronic cocaine administration during the development phase of behavioural sensitization. These differences show that by not capturing the data identified by the alternate search, we are missing a number of studies, which may be very similar in their setup, but are measuring different underlying biological concepts, perhaps of benefit to translation of results in future.

Evaluating the impact of missing these studies showed that overall effect of acute cocaine administration on locomotor activity was greater in data collected as part of the alternate search, when compared to data from the original search. In contrast, therapeutic agents administered seemed to be much more effective at improving behavioural outcome in data identified through the original search as opposed to the alternate search. Statistical heterogeneity observed in each of these two datasets was similar within acute studies.

For repeated cocaine paradigms there was less of a difference between overall effect sizes for data from the two different datasets, meaning the pooled effect size was not substantially different from that observed just from experiments identified in the original search. In terms of observed heterogeneity, in studies measuring behaviours after withdrawal from these chronic drug paradigms there was a reduction in heterogeneity seen in the alternate dataset compared to that seen in the original search dataset. When therapeutic agents were tested on the reversal of this behaviour, there was an opposite effect where observed statistical heterogeneity was much more substantial in the alternate search dataset, compared to that observed in the original search dataset.

Identifying these additional studies alongside those identified in my initial search, would have given me a larger sample of data leading to tighter confidence intervals and thus a smaller margin of error. Moreover, when looking at the effect of

therapeutic agents, more data means that any significant results that are seen can be considered more robust if repeated or generalized in different animal models using different animals, strains, sexes. Arguably, broad reviews such as this current one are at a threat from this kind of variability in reporting and labelling of studies. Moreover, studies using animal models of uniquely human disorders like psychiatric disorders, where any links between the model population and the human population with the disorder are mainly subjective, are also likely affected by difficulties in identifying all the data that is of relevance in the preclinical research field. For example, where the same outcome measure such as locomotor activity is used to describe a model for many different human behaviours in many different disease models, it can be argued that we might miss results that would be of value to a research question perhaps phrased differently to the original investigators.

7.3.4 Limitations

As mentioned resources only allowed for single data extraction. Depending on the error rate of data extractors this could mean that overall estimates of effect are prone to slight errors, however, any mistakes at this level are unlikely to have a substantial impact on the overall conclusions that are drawn. Data extractors were given continuous guidance throughout the extraction process with any issues discussed with a second investigator (myself). All data extractors followed a pre-defined protocol and 15% of both datasets were checked in detail by myself as the second screener. No substantial and recurring errors were encountered during this process and therefore the data were deemed robust. Unfortunately, due to time constraints the rest of the data extracted was only put to a sense check during analysis, where any missing data or any obvious outliers were double-checked and re-extracted if necessary. Clearly, issues like this call for more robust quality control, and advocate for double-data extraction. Of course, if time allowed in future, all datasets used for analyses above would ideally be double-screened to increase robustness of results.

We must also remember that conclusions based on data presented here are affected by the small sample of experiments, which affected my ability to perform multivariable meta-regression and assess for any co-linearity of variables. This could have also affected the overall estimation of publication bias. Moreover, studies collected as part of the original search were performed using a broad search and those identified in the alternate search were based on very narrow search criteria. It

has been shown by previous literature that results identified as part of a broad search and a more specific search can vary substantially, with the recommendation that both should be included in a systematic review (Egan et al., 2012). Finally, I only looked at the effect of updating the review of the field by a single year. For clinical systematic reviews there is a policy for relevance assessments every two years (Higgins and Green, 2008), however, there are no such guidance or criteria in place for animal studies and therefore it is unclear how often preclinical systematic reviews should be updated. It is difficult to say when these reviews become out of date as the rate of publication for animal studies seems much faster than for clinical studies, yet it is difficult to estimate differences in impact of new studies on overall conclusions. Finally, all of these analyses have been done on a subset of the literature identified during my initial systematic search and therefore might not be representative of the entire field.

8 Discussion

“Our field might become inundated with undigested data that collectively do not make sense”

(Tandon, 1999)

In 1999, a group of experts in the field suggested that perhaps part of our lack of understanding underlying schizophrenia was a result of the overwhelming amount of data that had been amassed in an effort to improve our knowledge of its biology in humans and how we might treat the disorder. The aetiology and the pathophysiology of the disorder were still elusive at the time and treatments available for those affected only provided moderate efficacy for managing the disorder, despite the vast amount of research in the field even then.

Arguably, little has changed in almost 20 years. Despite decades of research progress, our understanding of the disorder is still very limited, managing the disorder in the clinic still poses a challenge to psychiatry and developing entirely novel treatments have been largely unsuccessful. The drugs that have been developed in recent years build on similar mechanisms as more established antipsychotics, and include D2 receptor partial agonists like aripiprazole, brexpiprazole and cariprazine. These are sometimes called third generation antipsychotics, or dopamine system stabilizers (Mailman and Murthy, 2010) due to their novel and alternate mechanism of action compared to previously developed antipsychotics. Cariprazine has even been shown to be effective for negative symptoms (Debelle et al., 2015), and aripiprazole and brexpiprazole are approved as an adjunctive medication in the management of depression. Unfortunately, when placed in context with other existing antipsychotics evidence suggests that aripiprazole is not superior to other antipsychotics (Khanna et al., 2013) and scores only 9th on a list of fifteen antipsychotic drugs in terms of efficacy in schizophrenia (Leucht et al., 2013). Moreover, while these drugs have less metabolic and weight gain effects than other antipsychotics (Khanna et al., 2013), they are not without adverse side effects and a number of individuals do discontinue use due to these reasons (Frankel and Schwartz, 2017). A COCHRANE review of the efficacy of aripiprazole vs placebo in the treatment of schizophrenia, found that while aripiprazole has efficacy for schizophrenia, the data that shows this in clinical trials is of low or certainly questionable quality (Belgamwar and El-Sayeh, 2012).

This suggests that perhaps even these antipsychotics add little to the already long list of other antipsychotics being used in the clinic, and therefore research efforts into novel drugs of therapeutic value in schizophrenia continue.

A classic, coherent approach to development of novel treatments is challenging as our understanding of the aetiology of schizophrenia and the molecular mechanisms of action of current antipsychotics remains incomplete. Most therapeutic compounds approved for schizophrenia have been developed using hypotheses based on observations made from existing treatments. For example, aripiprazole was discovered using phenotypic screening of established targets to identify alternate molecular mechanisms of action for the reduction of dopaminergic overactivity (Swinney and Anthony, 2011). As a result, many therapeutic compounds are not entirely “novel” in this sense, but can offer perhaps improved solutions to issues seen in earlier drugs, such as a less severe side effect profile (Khanna et al., 2013)

The use of animal models for the study of schizophrenia and other related psychotic disorders has remained widespread, however, little of this information has been translated forward into clinical research (Moore, 2010). This arguably is somewhat of a far-removed domain of research especially for a uniquely human disorder such as schizophrenia and with the disorder being an undoubtedly complex one. It seems that we have something to gain from these experimental models; however, it is not always clear how much of this will have the power to ‘predict’ new and effective treatments for schizophrenia as well as other psychotic disorders. While some model paradigms are well established and discussed in the field as part of a number of excellent reviews (Jones et al., 2011; Marcotte et al., 2001), others are not as commonly described in the context of schizophrenia.

In the context of the large collection of research that has been published in the literature, it has become increasingly difficult to make sense of, perhaps for lack of a better word, ‘true’ models of psychotic disorders and understand the exact role they are able to play in improving our understanding of these disorders and their efficient management. In the context of these questions, my aim was to look at what could be affecting the translation of knowledge from the animal research field to subsequent domains of clinical research and practise.

In order to solve a problem, we first need to understand the problem, and therefore the aim of this project was to get a better understanding of the current state of the literature and quantitatively attempt to summarise aspects of this field through a systematic review and meta-analysis.

Of course, there are a number of general issues with summarising research, which also pertains to the schizophrenia research field. It can be difficult to summarise a field where there are a large number of studies all reporting discrete findings, which are rarely replicated. While some findings are corroborated by subsequent studies, the context in which to place new findings is often not clear.

In this review, I have shown in fact, that the literature available on this topic is vast, having identified and categorized almost 4000 studies of relevance. In order to better understand the preclinical field and its relevance to the human conditions it aims to model, I explored four main themes that I believe are building blocks of good pre-clinical research practice and thus are essential in informing subsequent research domains like clinical research (Figure 8.1). I argue that if compromised these components could explain failures in this information translation process. In this chapter, I review my main findings on what issues in the preclinical literature might be limiting translation and what considerations in future research could help to minimise these issues.

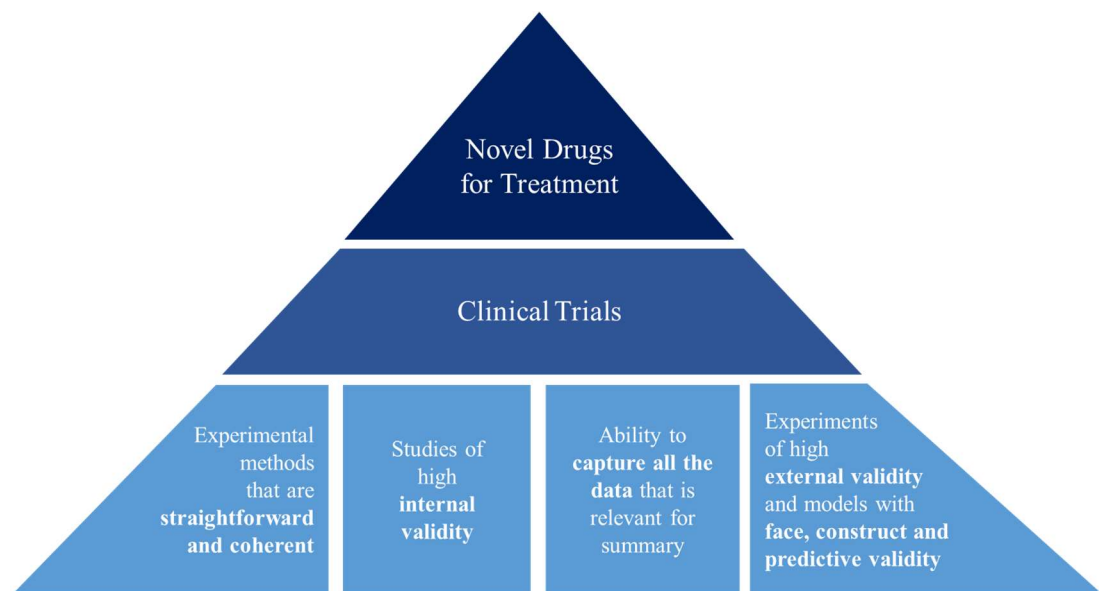


Figure 8.1 Research in basic science needs to be able to satisfy 4 key qualities of robust and reliable research in order to effectively support subsequent clinical research and practice

8.1 Models most often reported in the literature have limited relevance to psychotic disorders in the clinic

A resource compiled in 2014 recognised 149 different animal models used for research in schizophrenia, most models being genetic models (Koenig, 2014). Having performed a broad and shallow systematic review of the literature, here I have shown that this number is proportionally higher, in total having captured over 800 ways of inducing psychotic or schizophrenic-like behaviours in animals.

Clearly, the definition of an animal model of “schizophrenia” is subjective, however, all these publications claimed to have been looking at their research in relation to psychosis or more specifically schizophrenia in one way or another, or would have induced a model using methods otherwise regarded as a model in other publications. Some also claimed to have relevance to other psychiatric disorders, which makes it even more difficult to label an experimental setup in an animal as an ‘animal model of schizophrenia’.

A summary of the pre-clinical field of psychotic disorders shows that by large, animal models created using pharmacological interventions have continuously dominated the field. These models have been back-translated from clinical observations and have increased our knowledge substantially of underlying biology, by indicating dopaminergic, serotonergic, glutamatergic and other mechanisms in the psychopathology of schizophrenia (Moore, 2010).

For those publications claiming to model schizophrenia, the weaknesses of using drugs to induce a model, is their limited face validity and their lack of ability to recapitulate the neurodevelopmental nature of schizophrenia (Mattei et al., 2015), as well as their lack of validity in modelling aspects such as negative symptoms and cognitive deficits of schizophrenia (Jones et al., 2011). Schizophrenia in the clinic is chronic and episodic with different symptom domains appearing at different stages of the disorder (Steeds et al., 2015). It is also well established that there are likely to be multiple biological pathways at play as part of an integrative circuitry in the manifestation of the disorder (Benes, 2009). As most animal experimental paradigms using these pharmacological methods to induce their models involve acute administrations of these drugs and considering that these agents tend to be specific to single neurotransmitter systems, the ability of these models to present a complete picture of the clinical disorder is limited. Importantly, any treatment

compounds discovered through the reversal of the effects of these psychotic pharmacological agents, are constrained by the pharmacology of these agents to the specific mechanisms they are acting on (Moore, 2010), therefore limiting their utility in finding treatments with novel routes of action and treatments for negative and cognitive deficits.

In addition to a large number of pharmacological methods used to induce animal models, data collected here show that in the past couple of years there has also been a surge in the use of genetic models to study schizophrenia. Genetic models target a wide range of different genes to create so-called “risk factor models” (Moore, 2010). It is now well established that the genetic liability of schizophrenia is polygenic and genetic risk is conferred by a large number of alleles that exist in many different genes (Rees et al., 2015). As schizophrenia is associated with a large number of common genetic variants, all of which are rare and only contribute small effects of risk for the disorder (Nestler and Hyman, 2010), it is questionable how useful genetic models with single mutations can be. While GWAS studies have identified some drug targets such as genes involved in calcium channel signalling (Purcell et al., 2014), many of these common genetic variants have also been implicated in other psychiatric disorders. As a result, any models created using genetic manipulation of these targets would not only not be specific to schizophrenia, but would also not provide selective treatments for schizophrenia (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013). If genetic models continue to be used these should mainly address highly penetrant mutations (Nestler and Hyman, 2010). It is unarguable that in the clinic, the heterogeneity of the genetic makeup of different individuals is high. This implies that a treatment developed in a single gene knockout model is unlikely to help the entire patient population diagnosed with schizophrenia. This further limits the translatability of any results gained in these animal models to clinical research.

Biological pathways and affected genes might be studied in more reliable ways that do not require genetic manipulations or acute pharmacological interventions. For example, data presented here show that in recent years there has also been an increase in the reporting of developmentally induced animal models of schizophrenia. These models are based on human epidemiological observations and it is possible that environmental challenges during development in these models might initiate cell responses via epigenetic mechanisms as part of the animal's adaptation to a shift in external conditions which can induce similar effects as those seen in genetic models (Mattei et al., 2015). This is seen in animals born to mothers infected with Poly I:C during gestation that show naturally low levels of reelin expression, observed in schizophrenic patients, and the same effect as seen in animal models of schizophrenia where reelin is knocked out (Mattei et al., 2015).

8.2 Assessment of animal models in the field is disordered and inconsistent

As animals are not naturally psychotic or schizophrenic for that matter, both model and outcome measure used to characterize models need to be of relevance to provide an animal model of high fidelity for psychiatric conditions (Geyer, 2008). In total, I recorded 336 different behavioural outcome measures reported in the literature. These tests measured behaviours in animals that corresponded to about 34 different human symptoms or changes in natural behaviours. This clearly demonstrates that in the pre-clinical research field of psychotic disorders there is a substantial heterogeneity in the assessments used to measure behaviours in animal models by. Some tests in animals can be used to measure different behavioural domains and therefore sometimes the same test would be used to refer to measures of different behaviours.

At current, there are no strict rules for behavioural assessment in these animal models and the extent of their relevance to behaviours in schizophrenia. Some behavioural measures, however, can be tested in both rodents and humans such as pre-pulse inhibition of the startle response and cognitive tests defined as part of the Cambridge neuropsychological test automated battery (CANTAB) that can lead to robust associations between animal models and schizophrenic individuals (Geyer, 2008). Unfortunately, even pre-pulse inhibition can be modulated by various brain regions in animals and therefore its relevance to human schizophrenia can be

argued if it is entirely based on the argument that it involves the same brain regions in animals as it does in humans (Powell and Miyakawa, 2006).

Some obvious arguments against simple behavioural measurements in animals being representative of schizophrenic symptoms in humans are that they are too reductive and are measuring the use of brain structures that are too evolutionarily distant from complex human phenotypes. For this reason, especially, it is important that multiple tests are used to ascertain any one type of behavioural dimension (Powell and Miyakawa, 2006). There is no single behavioural task that captures the full clinical spectrum of schizophrenia that is diagnostic evidence in humans or is uniquely relevant to the disorder. Therefore, quantification of animals needs to employ a battery of tests, but the description of these tests I believe should also be as consistent as possible across different studies and different laboratories so meaningful comparisons can be made between them. This change would make it easier for novel research to build on previous knowledge. It is thought that adherence to established and common standards, for example when outcomes are universally agreed and are straightforward about what exactly they are measuring, is more likely to lead to true and meaningful findings (Ioannidis, 2005).

8.3 Lack of repetition to increase the robustness of results of individual studies

Ultimately, the data presented in this review shows that the literature is extremely broad in terms of exact methods used to induce a model, treatments tested and measures used to assess these models and treatments.

Not only is the field highly fragmented, making summarising it and putting new research into the context of others' difficult, but also studies are rarely replicated and therefore theories are not always corroborated. We see from this review that while a large number of potential treatment compounds have been tested across the literature many of these have only been considered in one or two studies. In fact, results in this review show that about 60% of treatments reported in the literature were only assessed in one study. It is possible that some of these compounds were not further explored due to lack of initially promising results, however, unsurprisingly many studies came to the conclusion that while their study showed important evidence for a phenomenon in our understanding of psychosis and schizophrenia, further research was needed to support their theory and take their data further.

While I did not assess the connection between different studies, it seems to me that there is little follow-up research leading on from previous research for some of the studies reviewed here. Replicability is imperative if we are to amass a collection of data reliable enough to drive and underpin theories towards better understanding and treatment of a disorder in the clinic and therefore I would suggest that the studies that are not used as building blocks for subsequent studies are wasteful. I believe a more systematic approach to research across different centres around the world would help to overcome this problem and maximise the potential of all findings.

In the clinic as mentioned above, multiple biological pathways are believed to be dysfunctional as part of an integrative circuitry and it is thought that any treatments that selectively only target one pathway are not likely to meet all the therapeutic needs of the disorder (Li et al., 2016). Many of the treatments assessed here were only explored in terms of their interactions with a single pathway. Different avenues of schizophrenia research have led to candidate drugs of more diverse mechanisms of action being tested in Phase II and III clinical trials (Geerts, 2016). Unfortunately, there has also been a high rate of false positives, as many drugs have shown efficacy in animal models, but have failed in clinical trials (Moore, 2010). Studying biological pathways in isolation can lead to novel insights into their working, but ultimately, we need more use of models and treatments in these models, which are able to interact with multiple pathways. Evidence shows that patient-reported quality of life is higher in schizophrenic individuals taking second-generation antipsychotics over those taking first-generation antipsychotics (Gründer et al., 2016), and it is thought that this superiority in some cases might be due to second-generation antipsychotics interacting with serotonin 5-HT_{2A} receptors as well as dopaminergic D₂ receptors (Mauri et al., 2014).

While involvement of more than a single pathway might yield treatments that are able to manage more complicated symptoms such as negative symptoms and cognitive deficits (Li et al., 2016), future studies need to have an important balance between the extent to which different pharmacological targets are engaged as it can lead to adverse side-effects. It has been suggested that by constraining animal models through etiological and pathogenic theories of schizophrenia based on clinical observations, we will also be able to reduce the number of false positive drugs being developed for treatment in humans (Moore, 2010).

8.4 Limitations in external validity of experiments

Animal models are used so that results collected in this sample can then be generalized to the target population of human sufferers of the disorder – the robustness of this causal relation and the extent to which this can be done is a measure of external validity (Belzung and Lemoine, 2011).

One of the most obvious and least surprising observations from this review was that most experiments measured behaviours of mostly male animals. The significance of this bias in the pre-clinical research field of psychotic disorders is that there are substantial differences between the sexes in both human schizophrenic individuals and animals, and in their basic response to experimental manipulations and measurements. Differences in humans include differences in incidence and the course of the disorder (Hayes et al., 2012), symptom severity (Ochoa et al., 2012), and differential response to antipsychotic treatment (Smith, 2010). Animals on the other hand show differential sensitivity to interventions used to induce models and perform differently in certain behavioural tests (Kokras and Dalla, 2014). While females are often omitted from studies for fear of hormonal cycles affecting the homogeneity of results and confusing the direct effects of experimental interventions (Beery and Zucker, 2011), this is the very reason why I think more research should focus on creating female animal models of schizophrenia and psychosis in general. Schizophrenia among other psychoses is thought to have a hormonal aetiological component (Hayes et al., 2012) and epidemiological data indicate that oestrogen may be protective in women (Riecher-Rössler, 2002). Women pre-menopause show a more benign course of schizophrenia, which changes after due to falling levels of oestrogen during menopause, and in turn, begins to require more severe treatment. As mentioned in Chapter 3 few studies focused their research specifically on puerperal psychosis and psychosis associated with the onset of menopause. For those affected by these conditions, this is a very real and debilitating problem (Crow, 2016; Robling et al., 2000) which unfortunately does not get enough distinct recognition in current classification systems (Rai et al., 2015). Research using female animals would increase our understanding of how to differentially treat schizophrenia in women, but also how to treat those women who suffer from psychosis in relation to hormonal fluctuations.

Second, there is little variability in species of animals most widely reported to be used in the literature. Mice used are most often C57BL/6 mice and rats used are mainly Sprague Dawley or Wistar rats. It is difficult to argue that results gained from a study that reports on the behaviour of an animal model measured in a single strain, which is then not replicated in any other strains, is generalizable to humans. In research, the argument for not using various strains is similar to that for not using both sexes: it is thought to introduce too much variability into behavioural results (Chadman et al., 2009). While it is argued in the literature that this genetic variability will reduce the likelihood of detecting a behavioural phenotype in relation to the manipulation of interest (Kannan et al., 2013), I would argue that variability can also be beneficial and essential in many cases. For example, genotype-phenotype relationships don't even necessarily generalize across different outbred strains of mice as shown by a study systematically looking at the phenotype of 30 inbred laboratory strains genetically mutated to model brain abnormalities showing differential and sometimes opposing responses (Sittig et al., 2016). Therefore, by studying a single genetic background, we are unlikely to make direct inferences, which are robust enough to generalize to humans. Perhaps one way to better model the genetic heterogeneity that we see in the clinic is to create and evaluate the same models as well as novel treatments tested using them, in various animal strains. This would maximise the spectrum of possible observations and thus increase the generalizability of results in this field of research to a larger proportion of those diagnosed with schizophrenia. Moreover, treatments that are shown to work in some strains, but not others, might also reduce the number of false positive drugs taken forward to clinical trials that end up failing to show efficacy in humans.

As schizophrenia is a neurodevelopmental disorder, the age at which models are investigated, especially developmental models, is an important variable that can affect the generalizability of preclinical findings to treatment regimens used in the clinic. In the clinic, psychotic symptoms in schizophrenia normally manifest themselves in adolescence and early adulthood in men and in slightly later in women. Developmental models allow for the unique opportunity to study disease progression over the developmental period in animals, potentially using models for the study of pre-clinical applications of preventative treatments that might be able to halt disease progression in humans. Data collected here shows that models were largely studied at adult stages of life and almost all treatments reported in experiments analysed were also tested in adult animals. It should be remembered,

however, when making inferences at these stages of life, different species show differences in the timing of key maturation events (Semple et al., 2013). Nevertheless, in recent years, there has been a lot of interest in the prevention of psychiatric illnesses through treatment during the prodromal phase of schizophrenia and in so-called “high-risk” individuals who have subtle symptoms, and some of whom later go on to develop full-blown schizophrenia (Mokhtari and Rajarethinam, 2013). While intervention at these early stages is still a controversial topic due to concerns over how much predictive validity criteria to identify these individuals actually have, research suggests that early intervention can potentially delay or prevent transition to psychotic disorders (Stafford et al., 2013). In order to improve our understanding of the feasibility of these treatment options and to increase our general understanding of the development of the disorder and its underlying biology, future animal studies of schizophrenia should consider animals at all stages of life. Finally, in this review, I used a standardized measure to categorize time of assessment of animals into specific stages of life in order to summarize studies easier. I did notice, however, that many studies reported fairly arbitrary age categories within their publications. In the literature, a recent study has highlighted that categories of rodent age can vary substantially between different laboratories, with animal researchers referring to rodents as an “adult” when aged anywhere between 6 to 20 weeks (Jackson et al., 2017). This lack of consistency in reporting can further hinder translatability of data collected in animal models and should be improved.

8.5 Limitations in face validity of developmental animal models

Face validity of animal models is a measure of phenomenological similarity between the model and the human condition, and the assessment of this concept is therefore complicated by exact definitions of the human condition. For example, according to one of the earlier descriptions of face validity, an animal model should not only resemble its clinical equivalent behaviourally, but these features should also be specific to the condition being modelled, co-exist in a subgroup of individuals diagnosed with the disorder and should not include features that are otherwise not seen in the clinic (Willner, 1984). Of course, these latter two restrictions make it difficult for any animal model to have strong face validity for schizophrenia.

This is not only because symptoms in the clinic are not necessarily specific to the disorder, but also the analogous behaviours of schizophrenia measured in preclinical research are widely used in the context of other disease models to measure sometimes different aspects of behaviour.

As the quality of predictions from animal models to humans depend heavily on the agreement between measures taken in animals in the laboratory and measures taken in humans in the clinic, animal models with high face validity for a clinical disorder should measure behaviours with a good degree of homology to measures used in clinical research (Geyer et al., 2012). If measures in animal models are not what is subsequently assessed in clinical trials then the extent to which these models are able to predict efficacy of drugs will be limited (Markou et al., 2009). It is clear from this review that some of the most commonly reported behavioural endpoints measured in the literature are of relevance to positive symptoms. While in animals, it is easy to perform these analogues of positive symptoms, it is arguable that the face validity of these models is weak. For example, the validity of locomotor activity as a measure of the clinical state of psychosis should be considered. Primarily, it is a non-specific behaviour in animals and can be interpreted in many different ways in the context of animal models of different disorders (Pratt et al., 2012). Secondly, in the clinic humans are not normally hyperactive, however, can often display stereotyped behaviours (Geyer and Moghaddam, 2002), but increased activity levels are more often seen as a core feature of bipolar disorder as opposed to schizophrenia (Perry et al., 2010). Thirdly, it is thought that an increase in locomotion in animals is a result of increased dopaminergic activity in mesolimbic and nigrostriatal dopamine pathways (van den Buuse, 2010) and therefore is widely used in the pre-clinical literature as a marker of stereotyped behaviours in the clinic such as psychotic agitation. However, in the clinic individuals affected by schizophrenia as well as those categorized as “high risk” individuals, show increased presynaptic dopamine functions in the associative striatum as opposed to the limbic striatum as we had previously thought. This misalignment in anatomy based on clinical evidence shows that back-translation of clinical findings has been slow to animal model studies (Kesby et al., 2018). Therefore, as our understanding of the underlying biology of psychosis increases, it becomes increasingly clear that some existing methods used to measure models might require some fine-tuning or that we need to look to measure models using alternative, perhaps more effective

translational approaches that are in line with neuroanatomical and biological features in schizophrenia.

Future experiments could improve translatability of outcome measures by either using tests in experimental animals which are as close to those used in humans (Eagle and Robbins, 2003) or measuring tests in humans which have more similarity to those measured in rodents (Shipman and Astur, 2008). They also might focus more on behavioural tests, which assess negative and cognitive deficits in animal models. Recent research shows that severe negative symptoms and cognitive deficits all correlate with poorer quality of life in schizophrenic individuals (Karow et al., 2014), and behavioural analogues of these symptoms might have more face validity than behaviours for positive symptoms (Kesby et al., 2018). As shown from this review, analogues of these human symptoms are much less widely reported in the literature than those thought to be of relevance to positive symptoms. This is limiting the introduction of novel drugs in schizophrenia, which currently do not exist for negative and cognitive deficits. Unfortunately, negative symptoms and cognitive deficits are not specific to schizophrenia and therefore measuring these behaviours in animal models of schizophrenia challenges classic definitions of face validity. Of course, it will be always difficult to claim an animal is only specific to schizophrenia if we are to do this, as clinically schizophrenia is a polythetic definition, whereby the same diagnosis can be made for patients with a different set of symptoms of differing severity, course and showing a variety of different outcomes (Owen, 2014). This substantially confounds the extent to which an animal experimental paradigm can be said to have an adequate level of face validity for schizophrenia as seen in the clinic.

As it has been suggested that each symptom domain should be addressed separately in schizophrenic individuals, future preclinical studies should consider testing their animal models of schizophrenia using comprehensive battery tests to maintain high face validity for schizophrenia, especially when assessing the efficacy of new treatments. In the clinic, functional outcome is measured through composite scores calculated from performance on questionnaires or test batteries (Goetghebeur and Swartz, 2016). Obviously, we are unable to model questionnaires in animals, but by matching this multi-criteria testing approach through test batteries, we could increase the reliability at which we could detect important biological changes in a sample and then generalize these findings to the target population.

8.6 Limitations in construct validity of developmental animal models

Construct validity measures the extent to which underlying pathological mechanisms in animal models and in humans diagnosed with the clinical condition are homologous (Wilson et al., 2010). Despite the fact that some animal models like developmentally induced animal models have strong face validity, true construct validity is hard to ascertain, as animal models of psychiatric disorders will always be limited to some extent on the premise that the clinical picture of these disorders is not entirely straightforward. Firstly, the related symptomology is categorised with a high degree of subjectivity (Geyer and Markou, 2000). Secondly, diagnostic categories are continuously being revised and do little for improving problems with the unreliability of psychiatric diagnoses (Aboraya et al., 2006). But most importantly, the construct validity of clinical diagnostic criteria themselves, defined as “the degree to which a diagnostic construct delineates a group of cases that share common underlying etiological and/or pathogenic processes” is arguably low for major psychiatric disorders (Owen et al., 2016). There is a lack of construct validity of current syndromic diagnostic approaches because our exact understanding of the aetiology and underlying pathogenic processes of these disorders is still very much limited.

Analysis of the literature herein accentuates some limitations in construct validity of developmentally induced animal models including that experimental design is highly heterogeneous even in studies using the same method of model induction, as most notable within this review for animal models created using prenatal infection. Variables differed across studies in terms of time of infection administration, the dose of substance administered and number of administrations. Perhaps this variability in preclinical methods stems from our uncertainty of the role these variables play in schizophrenia in the clinic, making our ability to judge the construct validity of each of these different approaches difficult. In the clinic maternal infection during pregnancy has been linked to adult schizophrenia, however, our understanding of whether there is a ‘sensitive period’ of development in the womb remains unclear (Khandaker et al., 2013). Some studies have implicated the first and second trimesters of pregnancy as critical periods of vulnerability to various infections (Selemon and Zecevic, 2015), however, it is not clear how these clinical data translate back down to animal studies mimicking these experimental

paradigms. Data presented here show that infections were administered to mothers at all stages of gestation. Translation between the two species is further complicated by the simple fact that in rodents, brain development carries on after birth while in humans this is done *in utero*, making inferences about the correct stages of development between the two species difficult (Mattei et al., 2015). A solution proposed for this in the literature is the use of mice of the *spiny* strain, which show complete brain development at birth (Ratnayake et al., 2012). None of the studies included and reviewed here reported using this strain of mice, but future studies could consider using strains such as this one, which closer models human brain development *in utero* for models where disruptions to early development are at the core of the model. While variation in experimental design is able to increase external validity (Maier et al., 2010), a lack of replication of experiments reporting different variables to existing experiments can also add further uncertainties to the field if these differences are not adequately highlighted. In this review, for example, it was difficult to draw conclusions about the impact of this variability in studies, as many experimental characteristics were not reported to be further replicated in the literature. Therefore I reiterate here my argument for an increased level of consistency in the performance and reporting of experiments. Any type of variability in the literature pertaining to experimental design needs to be more consistently and comprehensively described as well as replicated in ensuing experiments to allow conclusions to be meaningful. For this, use of existing guidelines such as the Animal Research: Reporting of In Vivo Experiments (ARRIVE) guidelines (Kilkenny et al., 2010b), Gold Standard Publication Checklist (Hooijmans et al., 2010a) and Landis (Landis et al., 2012) guidelines should be more widely used by researchers. Moreover, their use and full disclosure should be mandated more thoroughly by funders and journals to encourage scientists in the field. In addition to this, registration of preclinical experiments in a centralized database similar to those used for clinical trials will aid replication and transparency of experimental design used in studies. A more systematic approach like this can identify gaps in our knowledge based on where research has been done and where more research is required so that we can increase the chance of successful translation of preclinical research data into clinical research. Furthermore, research should aim to carry out tests using standardized procedures across different laboratories, to minimise the effects of unknown confounding factors. Evidence shows that data collected in experiments of a seemingly identical nature across different laboratories can yield very different

results (Wahlsten et al., 2003). A comprehensive reporting of the exact methodology used can reduce the impact of unknown confounding variables affecting results between studies in different laboratories. Limitations like this can also be a strong argument for the performance of more multicentre animal studies in future (www.multi-PART.org).

Ultimately, to improve translation in this field, there must be a bidirectional feedback loop between animal research and clinical research (Kesby et al., 2018). In the context that schizophrenia is a disorder characterized by gene-environment interactions, combined models perhaps have stronger construct validity for schizophrenia in the sense that they allow us to more closely mimic the predicted aetiology of this disorder as well as other psychotic disorders by looking at the effect of at least two disruptive “hits” (Karl and Arnold, 2014). Only 20% of the overall literature reviewed here explored the use of combination models, and only 12% of experiments testing the effects of a treatment in developmental animal models involved a second hit. While these models are still not able to model the entirety and the complexity of the disorder we see in humans, they are useful basic multifactorial models, which can help improve our understanding of the complexity of the interactions and mechanisms that underlie the disorder (Karl and Arnold, 2014). Moreover, some of these combined models might be of relevance to schizophrenic individuals with psychiatric comorbidities (Buckley et al., 2009).

8.7 Limitations in predictive validity of developmental animal models

Predictive validity is a measure of the similarity at which an animal model responds to medication used to treat the disorder in question when compared to how humans in the clinic respond to the same medication. The literature reviewed here shows that a similar hierarchy of treatment efficacy was seen in animal models as we observe in the clinic among patients (i.e. clozapine is superior to rest of treatments in terms of efficacy).

As most experiments involved looking at the effect of drugs already clinically established in developmentally-induced animal models, full predictive validity of these models remains to be established as so far no entirely novel clinically approved therapeutics have been developed using these models (Geyer et al., 2012). Unfortunately, it seems that quite a few novel drugs are compared to these

more established drugs. While it can be useful to compare the efficacy of a new treatment to a known treatment that already shows some efficacy in patients, models that are developed and validated through known treatment drugs, leave little room for the discovery of drugs of alternate chemical structures and mechanisms of action. Therefore, if our ultimate goal is to develop novel treatment options that treat symptoms that are not effectively addressed by current drugs then this is a limited approach (Geyer and Markou, 2000). What we are really lacking in the field today, and therefore something future studies could focus more on, is models with good predictive validity for cognitive deficits seen in schizophrenia as current antipsychotics do not seem to have a great efficacy on this cluster of symptoms in the clinic.

Ascertainment of the true predictive validity of developmentally induced animal models is difficult as there are many differences between how human patients usually utilise and respond during the course of treatment and how novel treatments of potential efficacy are assessed in animals. Firstly, in the clinic maintenance treatment of variable lengths is recommended for individuals experiencing psychotic episodes and usually, the treatment course is long term especially for those with more severe symptoms (Lally and MacCabe, 2015). Analysis of evidence presented in this review shows that often antipsychotics are administered acutely, which does not allow for long-term analysis of efficacy and therefore has limited relevance for clinical applications of the same drug. Moreover, evidence from the literature shows that treatment drugs are often tested in preclinical animal models at doses that are not representative of the clinical condition being modelled (Kapur et al., 2003). For example, single injections of haloperidol have been shown to create a different pharmacokinetic profile as the same regimen in humans does (Kapur et al., 2000). This suggests that any potential new treatment options need to be tested more systematically, using treatment regimens that more closely represent clinical therapeutic regimens. It is suggested that full dose-response relationship studies are still needed for potential treatment drugs in order to allow for appropriate inferences to be made between animal models and humans (Goetghebeur and Swartz, 2016). One way to better represent clinical treatment regimens in future preclinical studies could be through the use of receptor occupancy studies in rodents similar to those in positron emission tomography (PET) studies of humans. These could help to translate dose-response relationships in animals into a clinical context, however, will

also come with the limitations that results will be specific for species, strain, method of administration and type of drug used (Kapur et al., 2003).

Further to this, in the clinic, about 30% of those diagnosed with schizophrenia do not respond to currently available treatment options (Lally and MacCabe, 2015). In studies reviewed here only 15% of experiments in developmental animal models reported a negative effect of clinically established antipsychotics. Until we understand more about the heterogeneity of the clinical disorder it might be difficult to create animal models of translational value for all individuals who need treatment for a psychotic disorder. Future studies should increase some of the focus on animal models specific for antipsychotic drug-resistant schizophrenia. In the clinic clozapine is the first-line medication for individuals who don't respond to two or more other dopaminergic drugs (Lally et al., 2016). Therefore animal models that show improvement in response to clozapine, but not other antipsychotics in certain behavioural domains (Mouri et al., 2013) might be of relevance when developing new compounds of potential therapeutic value for treatment-resistant individuals. These limitations could partly explain why many drugs which work in animals, do not work in humans. By overcoming some of these limitations in preclinical experiments their relevance and their ability to predict clinical outcomes could potentially be increased.

Lastly, it is not only positive effects of treatments that should be explored in animal models but also any negative impact a therapeutic drug may have on other behavioural domains. There are a significant number of side effects associated with both typical and atypical antipsychotics (Lally and MacCabe, 2015). In the preclinical research field few measures of these side effects exist (i.e. catalepsy and vacuous chewing) (Gobira et al., 2013), and evidence collected here shows that they are rarely reported in the literature. This lack of information in the preclinical research means that we cannot ascertain for sure that animals do in fact respond to existing antipsychotics with enough homology as humans do. Moreover, in humans, side effects take time to manifest and therefore animal models might increase predictive validity by assessing novel therapeutic drugs and any behaviours indicative of potential side effects over a longer period of time than currently measured.

Ultimately, treatment of these disorders will always seem to require more than just pharmacological treatment in order to raise these patients' quality of life (Savilla et al., 2008). In the clinic, effective management of psychotic disorders often involves

some aspect of psychotherapy and social support (Owen et al., 2016). Pre-clinical tests looking at the effects of natural contextual changes like environmental enrichment could help us understand more about how non-pharmacological treatments might be able to support and possibly potentiate the effects of pharmaceutical treatments to give patients a more complete relief from the disorder.

8.8 Overall reporting of risk of bias and other methodological quality criteria is low

A failure in seeing drugs that show efficacy in animals then also showing efficacy in human clinical trials may also be in part attributed to inadequate design, conduct and reporting of animal experiments distorting overall results (Macleod et al., 2009). The reliability of results obtained in preclinical studies is highly dependent on the rigour at which an experiment is designed and carried out (Macleod et al., 2015). Systematic reviews of clinical trials show that inadequate methodological approaches are associated with bias (Jüni et al., 2001). Similarly, flaws in the quality and reporting of the methodology of animal studies can lead to systematic bias in an experiment and result in wrongful conclusions (van der Worp et al., 2010).

Based on the literature reviewed it is obvious that the reporting of aspects of experimental approach which might reduce the risk of bias and thus the likelihood of introduction of systematic errors into an experiment, is generally low across the psychotic disorders preclinical literature. These include a low percentage of reporting of random allocation of animals to group, blinded conduct of experiment and blinded assessment of outcome as well as a pre-defined explanation of inclusion and exclusion of animals. Additionally, one of the most poorly reported items is the reporting of the performance of sample size calculations. This overall was only reported in about 1% of the literature, which is concerning as we know that not performing a sample size calculation can lead to imprecision (Krauth et al., 2013). As animal studies often use very small group sizes in their experiments, unless these numbers are based on valid sample size calculations, they will be less precise, giving less accurate estimates of a true effect and will add very little to the research field (Button et al., 2013), arguably wasting resources and unnecessarily causing animal suffering in some cases. These findings are similar to those seen in other preclinical neuroscience domains (Macleod et al., 2015) and affect both independent study conclusions, and meaningful interpretations of evidence collected

in a field relating to a particular question. Moreover, the literature reviewed here and other fields of preclinical research (Hirst et al., 2014) reiterate the importance of future studies performing and reporting these experimental steps. They show that studies, which do not report randomisation, allocation concealment or blinded assessment of outcome report larger estimates of effect than those studies that do report these items. At its worst, this can lead to overestimation of the efficacy of a therapeutic drug in animals, which is then taken forward to and subsequently fails in clinical trials. To a lesser extent, this can mean that those studies, which are thus deemed unreliable, will have essentially been wasteful of resources used.

Study quality has been studied a lot more extensively in clinical trials compared to animal studies and therefore recommendations of good study quality for animal studies are largely based on the clinical CONSolidated Standards of Reporting Trials (CONSORT) statement (Altman et al., 2001). Arguably, however, different measures might be relevant for animal studies. Research collected here shows that the prevalence in reporting of these measures varies across the field depending on experimental design variables such as methods used to induce the model or measure outcome of the model in question. While in some cases it may not be appropriate to report these measures, it could also imply in other cases that investigators consider the impact of these measures to be less relevant for some experiments over others. As we move towards more advanced technological tools to measure behavioural and electrophysiological outcomes especially, some of these methodological considerations will no longer be relevant. For example, blinded assessment of outcome, in my opinion, is perhaps less relevant in studies where behaviour is measured using automated tools, such as beam breaks that record locomotion of animals to a computer. I would argue blinded assessment of outcome and the reporting of this concept becomes redundant in this scenario. Nevertheless, other important methodological criteria might come to the forefront in these cases, which we currently do not widely assess for (Krauth et al., 2013). Rosenthal's original study on experimenter bias, for example, also showed that an investigator's bias towards a group of animals affected the way they handled the animals – rats that were expected to perform better were handled with more care and friendliness (Rosenthal and Fode, 1963). Handling of animals is very rarely reported in the literature (Avey et al., 2016) as it is difficult to generalize not just between different institutions, but also individuals. Clearly, this is of less relevance in clinical trials and of more relevance to animal studies, where it is strongly argued that environmental

changes in laboratory environments can affect the behaviour of animals and thus have adverse effects on research data (Bailey, 2017; Gerdin et al., 2012). Considering schizophrenia has an established environmental nature and based on discussions with other animal researchers I believe that this is an important measure, which deserves more consideration and more detailed reporting in the methodology section of animal studies, especially those conducted in the context of psychotic disorders. Currently this and other confounding factors differing across different laboratories such as housing and husbandry (Avey et al., 2016), as well as factors more specific to experimental setup such as details about measuring apparatus used and exact timeline of experiments, are either rarely or inconsistently reported in the literature according to previous research (Egan et al., 2016). This makes the attention that we give to these factors, their standardization and subsequent reporting such an important concept. Unfortunately, things such as handling of animals is a vague concept in many ways, limiting the introduction of specific evaluation criteria for this measure, however, other aspects of methodology can and should be standardized in order to increase consistency and comparability of data collected from multiple sources.

My findings here support the hypothesis that efforts to develop new treatments for psychosis may have been hampered by the quality of supporting animal research. To improve the future field of pre-clinical psychosis research not only must the quality of studies improve, but reporting of this quality must also improve. Arguments for incomplete reporting because of limitations due to a lack of space are less valid for modern publications as many journals now allow for online publication of supplemental data. It is suggested that with risk of bias assessments becoming more prevalent in animal studies reporting of animal research will likely improve (Krauth et al., 2013).

While many of the studies identified here predate publication of the ARRIVE (Kilkenny et al., 2010b), Gold Standard Publication Checklist (Hooijmans et al., 2010a) and Landis (Landis et al., 2012) reporting guidelines these are cardinal aspects of good experimental design. Their importance (Fisher, 1937; Rosenthal and Fode, 1963), as well as lack of reporting has been recognised for many years and therefore this low prevalence in reporting is concerning. Perhaps even more concerning is the fact that data collected here shows that reporting of many of these items has improved little since these reporting guidelines have come into publication shown both by data collected here and elsewhere in the literature (Baker et al., 2014; Gulin et al., 2015).

It is clear that a shift in practice and comprehensive reporting of methodology will require more than just authors to change their practice. Reviewers, funders and journals will play a big role if we are to improve the quality for reporting of animal studies (Landis et al., 2012). For example, I have shown here that reporting criteria thought to affect the reported efficacy of therapeutic compounds especially (Hart et al., 2012), such as a statement of potential conflict of interest, were also infrequently reported in the literature. This and the reporting of compliance with animal welfare regulations, however, were the only two measures to have substantially improved in recent years, likely as a result of more stringent demands from journals. This shows that journals play a big role in the improvement of comprehensive reporting of methodology including methodological quality criteria. A recent study shows that a change in editorial policy by journals through mandating the completion of a checklist at the point of manuscript revision can improve the reporting of risks of bias for *in vivo* research (Macleod and The NPQIP Collaborative group, 2017). While the study also shows that there is further room for improvement, novel initiatives trying to tackle the problems of reproducibility and translatability of animal studies by attempting to minimise these issues early on in the experimental process might help. Through the introduction of guidelines aimed at helping investigators plan their experiments, such as PREPARE (Smith et al., 2017) or the Experimental Design Assistant commissioned by the NC3Rs (Percie du Sert et al., 2017), animal experiments might be better conducted to begin with.

8.9 Incomplete reporting of collected data

Reproducibility and translation of findings in preclinical research are not only confounded by incomplete reporting of methods, but also of incomplete reporting of results. Poor or no reporting of preclinical data, when carried out, is detrimental to the design of future studies as well as hindering these studies' ability to inform research in the clinical domain. An interesting perspective in the literature highlights just how important publishing of all studies and of all results is in the drug developmental process. Markou et al. (2009) suggest that there is a strong feed-forward loop in drug development, which always favours the forward progression of test compounds from preclinical studies to increasing phases of clinical trials in the presence of positive findings of efficacy. This feed-forward loop overlooks all other data that is otherwise negative or neutral and even when results of a compound are mixed in animal models, the predicted consensus will be that the compound is efficacious. This only becomes a real problem at later stages of the drug developmental process when a compound fails during clinical trials (Markou et al., 2009). This problem is clearly exacerbated by problems with publication bias and selective outcome reporting.

Publication bias has been shown in many other fields of pre-clinical research (Sena et al., 2010). Based on the subset of data used here for the estimation of publication bias in the literature, there is no significant evidence to show that animal studies of developmentally induced animal studies might be missing from the literature. This, however, could be a bigger problem if looking at the entire literature of animal models of psychotic disorders. Nevertheless, the data does show that there is substantial selective outcome reporting in the entire field, with almost 70% of publications analysed reporting some form of omission of result data. Incomplete reporting of results or favouring those of significance and positive results could lead to a skewed image of how good we think a model is at recapitulating symptoms of schizophrenia in animals and how efficacious a treatment compound is at attenuating effects in these animal models. Therefore, data collected here does imply that we are in fact getting a distorted image of all the data that has been collected to improve our understanding of schizophrenia. This problem is not unique to this field (Tsilidis et al., 2013) and suggests that universal changes could make a difference in multiple preclinical research fields. Ultimately, studies and individual experimental outcomes where model animals do not show significant deficits in

behaviour or where therapeutic drugs do not show efficacy in attenuating behavioural or other outcomes in model animals, need to be made available. This is so that we do not overestimate the overall efficacy of a therapeutic drug in pre-clinical studies and then have these candidate drugs fail later in clinical trials. For model characterising studies it is essential in improving the validity of a model and our understanding of its relevance to the clinical picture.

Moreover, as replicability of studies is not as common as arguably it should be, it is important that we have access to all raw data that has been collected through research so that this can be built upon by any subsequent research. Unfortunately, raw data are not something that are often disclosed or shared making *post hoc* calculations and comparisons by other scientists difficult when replicating studies or using data for meta-analyses for example. This affects not only the robustness of any conclusions that are drawn in such a way, but also makes an analysis of publications time-consuming. Novel initiatives encouraging the pre-registration of animal study protocols on www.preclinicaltrials.eu, similar to initiatives that have been used for almost two decades now in clinical trials will help reduce duplicates and overcome the risk of reporting bias. As shown by data collected here the reporting of an available protocol for a study is at current rare practice in the preclinical research field. This makes it difficult to tell whether published results of a completed study were part of the original protocol and if there were any that were not reported in the final study. Further to this, transparent publishing sites such as www.F1000research.com, allowing null results and other research outputs such as posters and slides to be published and for work to be openly reviewed by others in the field, have the potential to increase transparency in the field and allow for faster publication of all the data that has been collected. Open access to data is also being made more available through the use of data repositories, which should increasingly be used to publish all raw data collected that makes it citable and shareable and will overcome any issues with lack of space in a manuscript for reporting of all data (Hooijmans and Ritskes-Hoitinga, 2013). All of these platforms will hopefully be the beginning of a shift in this field of research towards more transparency in all fields of preclinical science and future studies reporting on animal models of psychosis should consider interacting more with these platforms. In order to increase the efficacy and robustness at which evidence can be summarised using meta-analyses to then inform further research and decisions, I would like to see more accurate and complete raw data being reported in the field. This would mean these analyses can

be more precise, leaving less room for over- or underestimations of biological phenomena, conclusions can be reached much faster without having to email authors, and automated tools could be developed to analyse data that is more structured.

8.10 Lack of consistency in reporting of studies which further complicates timely review of data

It is not only publication bias that can lead to an incomplete picture of a research field. The time that it takes to review studies will also inevitably mean that the snapshot of the available research evidence gained will be incomplete. Difficulty in identifying preclinical studies of relevance can also mean some publications are missed. I show here for example that by using alternate search words we can identify a comparable amount of additional studies also of relevance to a specific research question. While perhaps less of an issue in other research fields, I believe that especially in the field of complex and uniquely human disorders such as those seen in psychiatry (i.e. Where classification between symptoms and underlying pathophysiology are blurred even in the clinic), it can be difficult to capture all of the animal data that is of relevance to the clinical disorder whether reported to be of relevance to it or not. While it is useful to refer to relevant human disorders to a certain extent, this can also be less useful when we are looking at an experimental paradigm that is applicable to many different disorders and thus will not be captured as part of a search for a single disease entity. Recent literature suggests that there are no categorical distinctions in psychiatry and that in order to improve clinical outcomes we need to consider biologically similar subtypes that are not bound by phenotypic diagnosis according to current psychiatric nosology (Kapur et al., 2012). Moreover, it is thought by some researchers that psychosis should be seen more from the perspective of a spectrum model that is phenomenologically and temporally continuous and its expression is transdiagnostic and bi-directional with other mental health disorders (Guloksuz and van Os, 2018; McGrath et al., 2016). The debate over the validity of this view is not of importance here, however, it shows that studies reporting on animal models could take a similar approach. Studies could instead of defining their models by categorising them into separate entities based on a single or number of disorders they were modelling, instead report on phenotypes they were focusing on. This could lead to a more logical and comprehensive structure of the pre-clinical research field of psychosis. And I believe that this would make the

identification of patterns in the evidence collected easier to identify and could lead to a more straightforward picture of how preclinical data can be translated into clinical research. New initiatives such as the introduction of the endophenotype-oriented Research Domain Criteria (RDoC) system are being proposed as a new framework for defining animal models (i.e. according to endophenotypes instead of the old approach of complex diseases) (Anderzhanova et al., 2017). The fact that more recently developed antipsychotic medication such as aripiprazole, brexpiprazole and cariprazine are also approved for the treatment of other psychiatric disorders, such as bipolar disorder and major depressive disorder (Frankel and Schwartz, 2017), supports this idea. Perhaps focusing on doing research by looking at phenotypes that are of relevance to multiple psychiatric disorders would help in the discovery of 'novel' compounds to those already being used for schizophrenia.

In addition to this, I think there is also a more general issue with the contents of publications. More specifically, I think the detail of reporting of titles and abstracts for example as well as indexing of studies could be vastly improved. I observed during this review that some studies would fail to mention important aspects of experimental setup such as a full list of treatments tested or behavioural outcome measures used, for example. Difficulties with identifying studies of relevance has been widely described for qualitative research (Barroso et al., 2003), but I would argue this is not limited to only these types of research. Understandably, most abstracts are limited by word counts, as well as by their purpose to summarise a study in a way that is concise and interesting to the reader and will make them want to read the article further (Grant, 2013). Nevertheless, a lack of uniformity in reporting of and publishing of all data that has been collected through pre-clinical studies can affect those trying to summarise the literature relating to a specific area of interest and therefore these limitations in reporting can affect these reviews. For example, many systematic reviews perform electronic searches by searching titles and abstracts of studies in electronic databases (Sampson et al., 2006). To maximise the number of publications to be found of relevance especially during the performance of a broad search, which has been shown to yield results different to more specific searches shown both by evidence presented here and elsewhere in the literature (Egan et al., 2012), where possible, authors should list all interventions and treatments tested. This is not only in an effort to increase replicability, but also so that where experiments are similar despite being investigated and reported in

different research contexts, we may also learn from looking at research avenues that are directly parallel to ours.

For example, it is entirely likely that drugs of therapeutic value for some symptoms of complex disorders like schizophrenia already exist, but are being used in other lines of medicine. And in fact, the more our understanding grows of schizophrenia, the more it becomes apparent that not only the dopamine system is involved. Thus in order to treat other symptoms in schizophrenia like cognitive deficits, researchers have begun to repurpose drugs that have already been approved, which affect other systems (Yang et al., 2017). Cognitive enhancers spanning multiple targets, including ones that affect glutamatergic or cholinergic pathways, anti-dementia or anti-inflammatory agents, or even hormones like oxytocin and other compounds such as omega-3 fatty acids have all been tested for schizophrenia (Yang et al., 2017). Unfortunately, despite these compounds showing positive results in other conditions such as attention deficit disorders, schizotypal personality disorder, results are mixed and often inconclusive about the effectiveness of these drugs in schizophrenia (Harvey, 2009). It has been suggested that this could partly be due to the fact that most of the clinical trials assessing the efficacy of these compounds have done so in chronic schizophrenic individuals instead of focusing on individuals who are at an earlier phase of their disorder, when cognitive deficits begin to appear (Vreeker et al., 2015). In fact research shows that cognitive remediation is more effective for adults who are at earlier stages of their disorder (Corbera et al., 2017). Perhaps focusing on individuals who have increased vulnerability for the disorder might yield different results for cognitive enhancers (Vreeker et al., 2015). From a pharmacological aspect, the efficacy of these add-on cognitive enhancers in schizophrenic individuals might also be affected by the interference of simultaneous use of antipsychotic medications, exacerbated by our lack of understanding of the full spectrum of effects of current antipsychotics (Harvey and Bowie, 2012). Moreover, it is not clear if the same accepted doses for use in other conditions should be used in the treatment of cognitive deficits in schizophrenia and therefore dosing of these add-on treatments is likely to be an important variable. The interference from antipsychotic medication may especially affect dose-dependent side effects (Harvey and Bowie, 2012). In order to understand these interactions, when developing dosing standards using pre-clinical exploratory studies for cognitive enhancers, experimental setups should reflect this common clinical picture of simultaneous use of antipsychotics and cognitive enhancers (Harvey, 2009).

Nevertheless, by using research from other branches of science we might increase our pool of knowledge of compounds that might help alleviate these deficits in schizophrenia, which are otherwise also seen in a multitude of other psychiatric conditions.

Identifying all potential sources of relevant data also introduces another problem, of course. In the context of the large amounts of data that continue to be published in relation to specific areas of research, it becomes increasingly difficult to understand and process all of this data in a timely manner. This is important so that reliable decisions can be made for future studies and higher research domains can be reliably informed based on all the evidence that exists. As the research field is so broad-ranging within the psychosis field, it is hard to imagine how human investigators might keep up to date with all the research that is continuously going on. While systematic reviews can help, they are also limited by the time that it takes for reviewers to screen and process data and therefore can often lead to reviews which are considered to be out-of-date (Beller et al., 2013). I have shown here through updating a specific subset of the literature that overall conclusions may survive for a period of time even if the search does not include most recent data. Nevertheless, it is difficult to tell how long it would take for evidence published after a search to change results significantly in a broader sense of the field (Shojania et al., 2007). I believe I have also shown in Chapter 5 that there is potential in using computer-assisted techniques such as text mining to help with the processing of these data by automating a simple data extraction stage of the review process. The use of this tool and its expansion to the extraction of other types of data are limited by the format that this information is in and the variability of wording used to explain the same concept across different research fields (Bahor et al., 2017). This, as before, warrants a desire for standardization of reporting of publications. Machine learning is currently being fine-tuned and utilised for screening of publications for relevance (Bannach-Brown et al., 2018), however, we are still struggling with the extraction of certain types of data from these publications, which arguably is one of the most time-consuming and least reliable stages of the process (Jelicic Kadic et al., 2016). While this search was done prior to the availability of machine learning tools for citation screening, the manual screening results of this review are currently being used in the testing and validation of such systems for future reviews. Tools such as this would allow for faster initial screening of abstracts especially for larger projects like this review. If initiatives mentioned above were enforced and there was

an improvement in preclinical research publication in terms of consistency and completeness of reporting of methods and of results, then systematic reviews could be further automated.

8.11 Limitations

Firstly, this work is essentially observational and not experimental and should be considered as hypothesis generating only. Secondly, much of the data used for analyses was extracted by a single person, whereas to ascertain maximum reliability in results double data extraction would be the gold standard. Moreover, many of the analyses presented here as part of this project could only be done for part of the literature due to time constraints. While conclusions are thought to be representative of the entire dataset, data from studies not extracted might yield differential results. Here, I have demonstrated the use of an automated tool that could both reduce the time that it takes to perform a systematic review as well as provide a second screener for when resources do not allow for this. Clearly, this approach requires further development and ideally future research will build on this as well as other tools currently in development so that they can be adapted to systematic reviews of other fields of preclinical research and can thus be of use to many researchers.

The studies reviewed here were identified a number of years ago now and an update of the search might reveal new or different findings, the impact of which is likely to be minimal, however, not clear. Manually screened data from this project is being used for the development of machine learning techniques for the screening phase of the systematic review process, potentially making future updates of the project quicker. Categorization of studies could also be automated in future updates of the project, by using the lists of categories for model induction methods, behavioural outcome measures and treatments tested, put together here. These could be used to generate a string of regular expressions to be used in text mining of novel publications.

Moreover, I have shown here that the methodological quality of studies has some serious limitations and inadequacies and therefore conclusions based on this data might affect the precision and reliability of overall estimates of effect. The inclusion of only high-quality studies in this review of animal data, as suggested in clinical systematic reviews, was not appropriate at this stage due to the overwhelming number of animal studies, which do not report important methodological criteria.

Exploring the extent of limitations in reporting of these measures and looking at the impact of this in the literature to be able to make recommendations in future, was recognised to be a much more appropriate approach for this review.

Other general limitations of all systematic reviews possibly affecting results include the fact that I was only able to review the evidence and its impact as it was reported in publications, irrespective of what might have actually been done in the laboratory. Differences in reporting may have been confounded by unaccountable things such as differences in what different journals ask for to be reported, as well as restrictions on word counts (McCann et al., 2017). For reporting of closely related items such as measures taken to reduce risks of bias, it is also difficult to account for co-linearity. Unfortunately, due to small sample sizes, I was unable to use multi-variable meta-regression to investigate this further. In addition to this, I have also shown that the heterogeneity in the field is high and therefore in my meta-analysis I grouped studies together with sometimes vastly different characteristics as well as outcomes measured in different ways. This limits the usefulness of calculated overall estimates of effect, however, exploration of this heterogeneity, as done here where appropriate and possible, can help to understand any differences in efficacy that do exist between groups (McCann et al., 2017). Unfortunately, for many analyses reported here while heterogeneity could be detected and quantified, it was not always possible to look at the exact sources of this.

Finally, it is important to highlight that all research is susceptible to bias, including that of my own. For now, current assessment of risks of bias in preclinical systematic reviews and thus their quality is assessed using expert opinion instead of any reliable evidence-based methods (Mueller et al., 2014).

8.12 Conclusion

It is clear that there is a failure of translation from preclinical research to clinical research in the field of psychotic disorders. The aim of this project was to improve our understanding of current and past preclinical uses of animal models in the psychotic disorders research field. Here I have reviewed a number of potential underlying issues that could be having a negative effect on the effective translation of results, although it is most likely that all of these factors contribute to some extent to the problem.

At the moment, we are largely presented with information in the literature but have little knowledge of what came before the study including what the original design of it was, shown by a low 0.2% of studies that reported having a protocol available for their work. Due to incomplete reporting, we are also not always clear on what exactly happened during the study, evidenced by a lack of reporting of measures taken to reduce the risk of bias and incomplete descriptions of animals used including details about cohort, time of model induction and time of assessment to name a few. Moreover, heterogeneity in all aspects of experimental design and reporting in the literature is great, with minimal consistency across different studies. Variability should be encouraged in terms of animals used in research, including their sex, strain and age, as this increases the external validity of data collected. Variability in experimental design, however, needs to be more consistently and comprehensively described. Lastly, we are not always aware of what happens after a study has been performed. Selective outcome reporting is an obvious issue and data collection is further limited by that findings are very rarely reported as raw data. When these building blocks that make up good quality preclinical research are compromised, then how can we expect to effectively build on this structure (Figure 8.2)? All of the limitations mentioned here in the field are also frequently mentioned for clinical trials (Correll et al., 2011), so why are we measuring animal studies by a different scale?

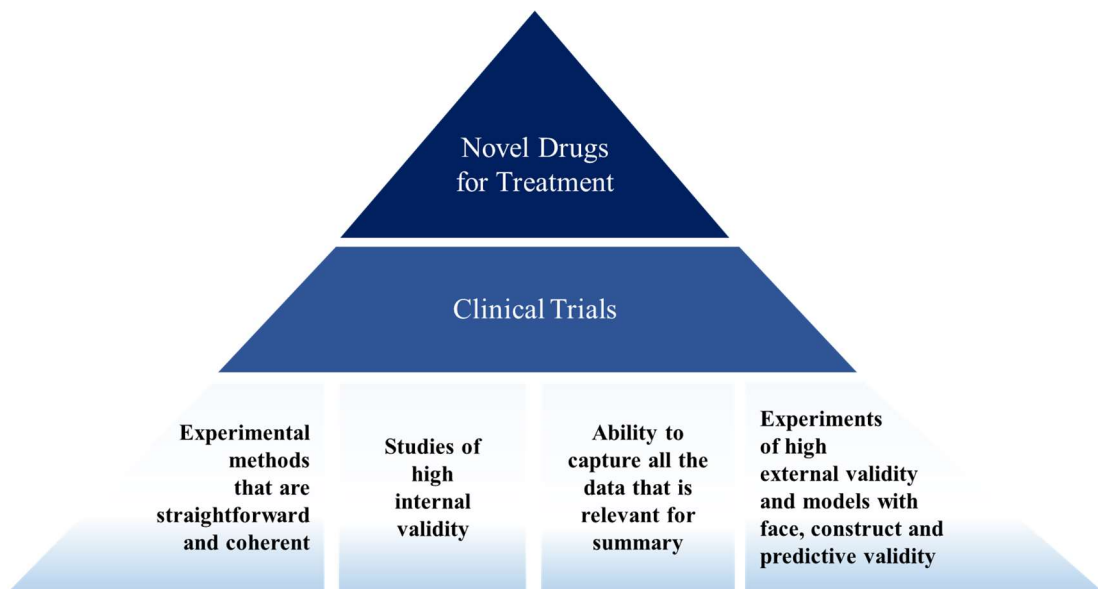


Figure 8.2 Translation of knowledge from basic science to clinical research and practice is clearly limited by four main issues, improvement of which might in turn also improve translation

It is important that future animal studies not only consider better experimental quality but also better reporting of said experimentation, including publishing of all data collected. Unless we are presented with a full and comprehensive picture, it is difficult to identify patterns in the literature, studies cannot be replicated and thus they lose their validity and reliability in informing subsequent domains of research. Reporting of all measures should be routinely done whether authors think that it might have a confounding effect on results or not. If a study is missing important information about experimental design then this also makes it difficult to compare and put into context with other studies. When we can't place animal work in the context of other work in the literature, has that piece of work added to our knowledge or has it been a waste of resources? I believe these are strong arguments for demanding a change in the pre-clinical research field of psychotic disorders.

It is also imperative that our methods of getting to important conclusions about the data are as fast as they can be, as reliable as they can be and as least wasteful or damaging to our resources as they can be. Undoubtedly collaboration is essential and that's why international consortiums such as Novel Methods leading to New Medications in Depression and Schizophrenia (NEWMEDS), a collaborative project between researchers in academia and industry for the development of new drugs in schizophrenia and depression, will be vital in agreeing on standardized methods in preclinical research that will be analogous to clinical research (NEWMEDS-europe.com). The price of pre-clinical research is high, but the price of testing drugs in clinical trials, which then go onto fail, is even higher. While there are a number of promising drug candidates in final stages of clinical trials at the moment (Fellner, 2017), whether they will make it into clinical use and how they will be received remains to be seen. In the meantime, animal models should continue to be a unique tool used in improving our understanding of complex psychiatric conditions such as psychosis spectrum disorders. The extent of their ability to inform clinical knowledge is not known, and will only become clear once we start doing and reporting research properly and thoroughly. The only thing that is certain is that there is a need for better treatment options for individuals who battle with psychotic disorders, and this certainly warrants the use and further development of tools, which might increase our knowledge of how to reach these end goals. I believe that in order to overcome the issue of having the research field of psychotic disorders "become inundated with undigested data that collectively do not make sense" (Tandon, 1999), improve translation of evidence between pre-clinical and clinical studies, and thus propel

preclinical psychosis research in the right direction, we must have a transparent, replicable and systematic approach to future research. Only by learning from and building on previous research in this way through systematic reviews will we truly be able to stand on the shoulders of giants and see further than is individually possible.

Bibliography

- Abi-Dargham, A., Rodenhiser, J., Printz, D., Zea-Ponce, Y., Gil, R., Kegeles, L.S., Weiss, R., Cooper, T.B., Mann, J.J., Van Heertum, R.L., Gorman, J.M., Laruelle, M., 2000. Increased baseline occupancy of D2 receptors by dopamine in schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* 97, 8104–9.
- Aboraya, A., Rankin, E., France, C., El-Missiry, A., John, C., 2006. The Reliability of Psychiatric Diagnosis Revisited: The Clinician's Guide to Improve the Reliability of Psychiatric Diagnosis. *Psychiatry (Edgmont)*. 3, 41–50.
- Adobe Systems Incorporated, 2006. PDF Reference, sixth edition: Adobe Portable Document Format version 1.7.
- Aghajanian, G.K., Marek, G.J., 2000. Serotonin model of schizophrenia: emerging role of glutamate mechanisms. *Brain Res. Brain Res. Rev.* 31, 302–12.
- Ahn, K., Gotay, N., Andersen, T.M., Anvari, A.A., Gochman, P., Lee, Y., Sanders, S., Guha, S., Darvasi, A., Glessner, J.T., Hakonarson, H., Lencz, T., State, M.W., Shugart, Y.Y., Rapoport, J.L., 2014. High rate of disease-related copy number variations in childhood onset schizophrenia. *Mol. Psychiatry* 19, 568–572. doi:10.1038/mp.2013.59
- Allen, K.M., Fung, S.J., Shannon Weickert, C., 2016. Cell proliferation is reduced in the hippocampus in schizophrenia. *Aust. New Zeal. J. Psychiatry* 50, 473–480. doi:10.1177/0004867415589793
- Altman, D.G., Schulz, K.F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gøtzsche, P.C., Lang, T., CONSORT GROUP (Consolidated Standards of Reporting Trials), 2001. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann. Intern. Med.* 134, 663–94.
- American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders : DSM-5. American Psychiatric Association, Arlington, VA.
- Ananiadou, S., McNaught, J., 2006. Text mining for biology and biomedicine. Artech House.
- Anderson, J.E., Wible, C.G., McCarley, R.W., Jakab, M., Kasai, K., Shenton, M.E., 2002. An MRI study of temporal lobe abnormalities and negative symptoms in chronic schizophrenia. *Schizophr. Res.* 58, 123–34.
- Anderzhanova, E., Kirmeier, T., Wotjak, C.T., 2017. Animal models in psychiatric research: The RDoC system as a new framework for endophenotype-oriented translational neuroscience. *Neurobiol. Stress* 7, 47–56. doi:10.1016/J.YNSTR.2017.03.003
- Andreasen, N.C., 1995. Symptoms, signs, and diagnosis of schizophrenia. *Lancet (London, England)* 346, 477–81.
- Aquila, R., Citrome, L., 2015. Cognitive impairment in schizophrenia: the great unmet need. *CNS Spectr.* 20, 32–40. doi:10.1017/S109285291500070X
- Arciniegas, D.B., 2015. Psychosis. *Continuum (Minneap. Minn.)*. 21, 715–736. doi:10.1212/01.CON.0000466662.89908.e7

- Avey, M.T., Moher, D., Sullivan, K.J., Fergusson, D., Griffin, G., Grimshaw, J.M., Hutton, B., Lalu, M.M., Macleod, M., Marshall, J., Mei, S.H.J., Rudnicki, M., Stewart, D.J., Turgeon, A.F., McIntyre, L., Group, C.C.C.T.B., 2016. The Devil Is in the Details: Incomplete Reporting in Preclinical Animal Research. *PLoS One* 11, e0166733. doi:10.1371/journal.pone.0166733
- Bahor, Z., Liao, J., Macleod, M.R., Bannach-Brown, A., McCann, S.K., Wever, K.E., Thomas, J., Ottavi, T., Howells, D.W., Rice, A., Ananiadou, S., Sena, E., 2017. Risk of bias reporting in the recent animal focal cerebral ischaemia literature. *Clin. Sci.* 131, 2525–2532. doi:10.1042/CS20160722
- Bahor, Z., Nunes-Fonseca, C., Thomson, L.D.G., Sena, E.S., Macleod, M.R., 2016. Improving our understanding of the *in vivo* modelling of psychotic disorders: A protocol for a systematic review and meta-analysis. *Evidence-based Preclin. Med.* 3, e00022. doi:10.1002/ebm2.22
- Bailey, J., 2017. Does the stress inherent to laboratory life and experimentation on animals adversely affect research data? *Altern. Lab. Anim.* 45, 299–301.
- Bailey, K.R., Crawley, J.N., 2009. *Anxiety-Related Behaviors in Mice, Methods of Behavior Analysis in Neuroscience*. CRC Press/Taylor & Francis.
- Baker, D., Lidster, K., Sottomayor, A., Amor, S., 2014. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PLoS Biol.* 12, e1001756. doi:10.1371/journal.pbio.1001756
- Bannach-Brown, A., Przybyła, P., Thomas, J., Rice, A.S.C., Ananiadou, S., Liao, J., Macleod, M.R., 2018. The use of text-mining and machine learning algorithms in systematic reviews: reducing workload in preclinical biomedical sciences and reducing human screening error. *bioRxiv* 255760. doi:10.1101/255760
- Barch, D.M., Sheffield, J.M., 2014. Cognitive impairments in psychotic disorders: common mechanisms and measurement. *World Psychiatry* 13, 224–32. doi:10.1002/wps.20145
- Barchard, K.A., Pace, L.A., 2011. Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Comput. Human Behav.* 27, 1834–1839. doi:10.1016/j.chb.2011.04.004
- Barnes, S.A., Der-Avakian, A., Markou, A., 2014. Anhedonia, avolition, and anticipatory deficits: assessments in animals with relevance to the negative symptoms of schizophrenia. *Eur. Neuropsychopharmacol.* 24, 744–58. doi:10.1016/j.euroneuro.2013.10.001
- Barroso, J., Gollop, C.J., Sandelowski, M., Meynell, J., Pearce, P.F., Collins, L.J., 2003. The Challenges of Searching for and Retrieving Qualitative Studies. *West. J. Nurs. Res.* 25, 153–178. doi:10.1177/0193945902250034
- Bartlett, E., Hallin, A., Chapman, B., Angrist, B., 1997. Selective Sensitization to the Psychosis-Inducing Effects of Cocaine: A Possible Marker for Addiction Relapse Vulnerability? *Neuropsychopharmacology* 16, 77–82. doi:10.1016/S0893-133X(96)00164-9
- Bastian, H., Glasziou, P., Chalmers, I., Cochrane, A., Duncan, A., Cummings, M., Barron, B., Bukantz, S., Banta, D., Stjernswärd, J., Chalmers, T., Chalmers, I.,

- Goldschmidt, P., Mulrow, C., L'Abbé, K., Detsky, A., O'Rourke, K., Sacks, H., Berrier, J., Reitman, D., Ancona-Berk, V., Chalmers, T., Oxman, A., Guyatt, G., Jenicek, M., Chalmers, I., Starr, M., Chalmers, I., Clarke, M., Oxman, A., Hopewell, S., Loudon, K., Clarke, M., Oxman, A., Dickersin, K., Lee, K., Bacchetti, P., Sim, I., Kass, E., Chan, A., Altman, D., Zarin, D., Ide, N., Tse, T., Harlan, W., West, J., Dijkers, M., Shojania, K., Sampson, M., Ansari, M., Coucette, S., Moher, D., Mallett, S., Clarke, M., Moher, D., Tetzlaff, J., Tricco, A., Sampson, M., Altman, D., Chalmers, I., Glasziou, P., Koch, G., Garritty, C., Tsertsvadze, A., Tricco, A., Sampson, M., Moher, D., 2010. Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up? *PLoS Med.* 7, e1000326. doi:10.1371/journal.pmed.1000326
- Battistella, G., Fornari, E., Annoni, J.-M., Chtioui, H., Dao, K., Fabritius, M., Favrat, B., Mall, J.-F., Maeder, P., Giroud, C., 2014. Long-term effects of cannabis on brain structure. *Neuropsychopharmacology* 39, 2041–8. doi:10.1038/npp.2014.67
- Bebarta, V., Luyten, D., Heard, K., 2003. Emergency medicine animal research: does use of randomization and blinding affect the results? *Acad. Emerg. Med.* 10, 684–7.
- Becker, J.B., Hu, M., 2008. Sex differences in drug abuse. *Front. Neuroendocrinol.* 29, 36–47. doi:10.1016/j.yfrne.2007.07.003
- Becker, J.B., Molenda, H., Hummer, D.L., 2001. Gender differences in the behavioral responses to cocaine and amphetamine. Implications for mechanisms mediating gender differences in drug abuse. *Ann. N. Y. Acad. Sci.* 937, 172–87.
- Beer, M.D., 1996. Psychosis: A history of the concept. *YCOMP</cjid> Compr. Psychiatry* 37, 273–291.
- Beery, A.K., Zucker, I., 2011. Sex bias in neuroscience and biomedical research. *Neurosci. Biobehav. Rev.* 35, 565–72. doi:10.1016/j.neubiorev.2010.07.002
- Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K.F., Simel, D., Stroup, D.F., 1996. Improving the Quality of Reporting of Randomized Controlled Trials. *JAMA* 276, 637. doi:10.1001/jama.1996.03540080059030
- Begley, C.G., Ellis, L.M., 2012. Raise standards for preclinical cancer research. *Nature* 483, 531–533. doi:10.1038/483531a
- Belgamwar, R.B., El-Sayeh, H.G.G., 2012. Aripiprazole vs Placebo for Schizophrenia. *Schizophr. Bull.* 38, 382–383. doi:10.1093/schbul/sbs044
- Bell, D.S., 1973. The Experimental Reproduction of Amphetamine Psychosis. *Arch. Gen. Psychiatry* 29, 35. doi:10.1001/archpsyc.1973.04200010020003
- Beller, E.M., Chen, J.K.-H., Wang, U.L.-H., Glasziou, P.P., 2013. Are systematic reviews up-to-date at the time of publication? *Syst. Rev.* 2, 36. doi:10.1186/2046-4053-2-36
- Belzung, C., Lemoine, M., 2011. Criteria of validity for animal models of psychiatric disorders: focus on anxiety disorders and depression. *Biol. Mood Anxiety*

Disord. 1, 9. doi:10.1186/2045-5380-1-9

- Benes, F.M., 2009. Neural Circuitry Models of Schizophrenia: Is it Dopamine, GABA, Glutamate, or Something Else? *Biol. Psychiatry* 65, 1003–1005. doi:10.1016/J.BIOPSYCH.2009.04.006
- Bitanhirwe, B.K., Peleg-Raibstein, D., Mouttet, F., Feldon, J., Meyer, U., 2010. Late Prenatal Immune Activation in Mice Leads to Behavioral and Neurochemical Abnormalities Relevant to the Negative Symptoms of Schizophrenia. *Neuropsychopharmacology* 35, 2462–2478. doi:10.1038/npp.2010.129
- Bobes, J., Garcia-Portilla, M.P., Bascaran, M.T., Saiz, P.A., Bousoño, M., 2007. Quality of life in schizophrenic patients. *Dialogues Clin. Neurosci.* 9, 215–26.
- Bracken, M.B., 2009. Why animal studies are often poor predictors of human reactions to exposure. *J. R. Soc. Med.* 102, 120–2. doi:10.1258/jrsm.2008.08k033
- Brady, K.T., Lydiard, R.B., Malcolm, R., Ballenger, J.C., 1991. Cocaine-induced psychosis. *J. Clin. Psychiatry* 52, 509–12.
- Braff, D.L., Geyer, M.A., Swerdlow, N.R., 2001. Human studies of prepulse inhibition of startle: normal subjects, patient groups, and pharmacological studies. *Psychopharmacology (Berl)*. 156, 234–58.
- Bramness, J.G., Gundersen, Ø.H., Guterstam, J., Rognli, E.B., Konstenius, M., Løberg, E.-M., Medhus, S., Tanum, L., Franck, J., 2012. Amphetamine-induced psychosis—a separate diagnostic entity or primary psychosis triggered in the vulnerable? *BMC Psychiatry* 12, 221. doi:10.1186/1471-244X-12-221
- Breier, A., Su, T.P., Saunders, R., Carson, R.E., Kolachana, B.S., de Bartolomeis, A., Weinberger, D.R., Weisenfeld, N., Malhotra, A.K., Eckelman, W.C., Pickar, D., 1997. Schizophrenia is associated with elevated amphetamine-induced synaptic dopamine concentrations: evidence from a novel positron emission tomography method. *Proc. Natl. Acad. Sci. U. S. A.* 94, 2569–74.
- Brickman, A.M., Buchsbaum, M.S., Shihabuddin, L., Byne, W., Newmark, R.E., Brand, J., Ahmed, S., Mitelman, S.A., Hazlett, E.A., 2004. Thalamus size and outcome in schizophrenia. *Schizophr. Res.* 71, 473–84. doi:10.1016/j.schres.2004.03.011
- Brisch, R., Saniotis, A., Wolf, R., Bielau, H., Bernstein, H.-G., Steiner, J., Bogerts, B., Braun, K., Braun, A.K., Jankowski, Z., Kumaratilake, J., Kumaritlake, J., Henneberg, M., Gos, T., 2014. The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: old fashioned, but still in vogue. *Front. psychiatry* 5, 47. doi:10.3389/fpsy.2014.00047
- Brody, A.L., Olmstead, R.E., London, E.D., Farahi, J., Meyer, J.H., Grossman, P., Lee, G.S., Huang, J., Hahn, E.L., Mandelkern, M.A., 2004. Smoking-Induced Ventral Striatum Dopamine Release. *Am. J. Psychiatry* 161, 1211–1218. doi:10.1176/appi.ajp.161.7.1211
- Brown, A.S., Patterson, P.H., 2011. Maternal infection and schizophrenia: implications for prevention. *Schizophr. Bull.* 37, 284–90. doi:10.1093/schbul/sbq146

- Brown, T.A., Barlow, D.H., 2005. Dimensional versus categorical classification of mental disorders in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders and beyond: comment on the special section. *J. Abnorm. Psychol.* 114, 551–6. doi:10.1037/0021-843X.114.4.551
- Buckley, P.F., Miller, B.J., Lehrer, D.S., Castle, D.J., 2009. Psychiatric comorbidities and schizophrenia. *Schizophr. Bull.* 35, 383–402. doi:10.1093/schbul/sbn135
- Burgy, M., 2008. The Concept of Psychosis: Historical and Phenomenological Aspects. *Schizophr. Bull.* 34, 1200–1210. doi:10.1093/schbul/sbm136
- Buscemi, N., Hartling, L., Vandermeer, B., Tjosvold, L., Klassen, T.P., 2006. Single data extraction generated more errors than double data extraction in systematic reviews. *J. Clin. Epidemiol.* 59, 697–703. doi:10.1016/j.jclinepi.2005.11.010
- Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–76. doi:10.1038/nrn3475
- Cadenhead, K.S., 2011. Startle reactivity and prepulse inhibition in prodromal and early psychosis: effects of age, antipsychotics, tobacco and cannabis in a vulnerable population. *Psychiatry Res.* 188, 208–16. doi:10.1016/j.psychres.2011.04.011
- Cadenhead, K.S., Swerdlow, N.R., Shafer, K.M., Diaz, M., Braff, D.L., 2000. Modulation of the Startle Response and Startle Laterality in Relatives of Schizophrenic Patients and in Subjects With Schizotypal Personality Disorder: Evidence of Inhibitory Deficits. *Am. J. Psychiatry* 157, 1660–1668. doi:10.1176/appi.ajp.157.10.1660
- Cannon, T.D., Chung, Y., He, G., Sun, D., Jacobson, A., van Erp, T.G.M., McEwen, S., Addington, J., Bearden, C.E., Cadenhead, K., Cornblatt, B., Mathalon, D.H., McGlashan, T., Perkins, D., Jeffries, C., Seidman, L.J., Tsuang, M., Walker, E., Woods, S.W., Heinssen, R., 2015. Progressive Reduction in Cortical Thickness as Psychosis Develops: A Multisite Longitudinal Neuroimaging Study of Youth at Elevated Clinical Risk. *Biol. Psychiatry* 77, 147–157. doi:10.1016/j.biopsych.2014.05.023
- Caporaso, J.G., 2009. Extracting signal from noise in biological data : evaluations and applications of text mining and sequence coevolution.
- Cardno, A.G., Marshall, E.J., Coid, B., Macdonald, A.M., Ribchester, T.R., Davies, N.J., Venturi, P., Jones, L.A., Lewis, S.W., Sham, P.C., Gottesman, I.I., Farmer, A.E., McGuffin, P., Reveley, A.M., Murray, R.M., 1999. Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch. Gen. Psychiatry* 56, 162–8.
- Carr, C.P., Martins, C.M.S., Stingel, A.M., Lemgruber, V.B., Juruena, M.F., 2013. The Role of Early Life Stress in Adult Psychiatric Disorders. *J. Nerv. Ment. Dis.* 201, 1007–1020. doi:10.1097/NMD.0000000000000049
- Carter, C.S., Barch, D.M., Buchanan, R.W., Bullmore, E., Krystal, J.H., Cohen, J., Geyer, M., Green, M., Nuechterlein, K.H., Robbins, T., Silverstein, S., Smith,

- E.E., Strauss, M., Wykes, T., Heinssen, R., 2008. Identifying Cognitive Mechanisms Targeted for Treatment Development in Schizophrenia: An Overview of the First Meeting of the Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia Initiative. *Biol. Psychiatry* 64, 4–10. doi:10.1016/j.biopsych.2008.03.020
- Casey, B.J., Glatt, C.E., Lee, F.S., 2015. Treating the Developing versus Developed Brain: Translating Preclinical Mouse and Human Studies. *Neuron* 86, 1358–1368. doi:10.1016/j.neuron.2015.05.020
- Celada, P., Bortolozzi, A., Artigas, F., 2013. Serotonin 5-HT_{1A} Receptors as Targets for Agents to Treat Psychiatric Disorders: Rationale and Current Status of Research. *CNS Drugs* 27, 703–716. doi:10.1007/s40263-013-0071-0
- Chadman, K.K., Yang, M., Crawley, J.N., 2009. Criteria for validating mouse models of psychiatric diseases. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 150B, 1–11. doi:10.1002/ajmg.b.30777
- Chalmers, T.C., Smith, H., Blackburn, B., Silverman, B., Schroeder, B., Reitman, D., Ambroz, A., 1981. A method for assessing the quality of a randomized control trial. *Control. Clin. Trials* 2, 31–49.
- Check Hayden, E., 2010. Sex bias blights drug studies. *Nature* 464, 332–333. doi:10.1038/464332b
- Connell, P.H., 1990. Amphetamine psychosis (Reply). *Br. J. Addict.* 85, 1505–1506.
- Corbera, S., Wexler, B.E., Poltorak, A., Thime, W.R., Kurtz, M.M., 2017. Cognitive remediation for adults with schizophrenia: Does age matter? *Psychiatry Res.* 247, 21–27. doi:10.1016/j.psychres.2016.10.084
- Corcoran, C., Mujica-Parodi, L., Yale, S., Leitman, D., Malaspina, D., 2002. Could stress cause psychosis in individuals vulnerable to schizophrenia? *CNS Spectr.* 7, 33–8, 41–2.
- Correll, C.U., Kane, J.M., 2014. Schizophrenia: Mechanism of Action of Current and Novel Treatments. *J. Clin. Psychiatry* 75, 347–348. doi:10.4088/JCP.13078co8c
- Correll, C.U., Kishimoto, T., Kane, J.M., 2011. Randomized controlled trials in schizophrenia: opportunities, limitations, and trial design alternatives. *Dialogues Clin. Neurosci.* 13, 155–72.
- Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 381, 1371–1379. doi:10.1016/S0140-6736(12)62129-1
- Crow, E.A., 2016. Schizophrenia during Menopausal Transition. *Ment. Health Fam. Med.* 12, 190–195.
- Curran, H.V., D'Souza, D.C., Robbins, T.W., Fletcher, P., 2009. Modelling psychosis. *Psychopharmacology (Berl.)* 206, 513–514. doi:10.1007/s00213-009-1663-8
- Currie, G.L., Delaney, A., Bennett, M.I., Dickenson, A.H., Egan, K.J., Vesterinen, H.M., Sena, E.S., Macleod, M.R., Colvin, L.A., Fallon, M.T., 2013. Animal

models of bone cancer pain: Systematic review and meta-analyses. *Pain* null.
doi:10.1016/j.pain.2013.02.033

- da Frota Ribeiro, C.M., Sanacora, G., Hoffman, R., Ostroff, R., 2016. The Use of Ketamine for the Treatment of Depression in the Context of Psychotic Symptoms. *Biol. Psychiatry* 79, e65–e66. doi:10.1016/j.biopsych.2015.05.016
- Davidson, L.L., Heinrichs, R.W., 2003. Quantification of frontal and temporal lobe brain-imaging findings in schizophrenia: a meta-analysis. *Psychiatry Res.* 122, 69–87.
- Davidson, M., Galderisi, S., Weiser, M., Werbeloff, N., Fleischhacker, W.W., Keefe, R.S., Boter, H., Keet, I.P.M., Prelipceanu, D., Rybakowski, J.K., Libiger, J., Hummer, M., Dollfus, S., López-Ibor, J.J., Hranov, L.G., Gaebel, W., Peuskens, J., Lindfors, N., Riecher-Rössler, A., Kahn, R.S., 2009. Cognitive Effects of Antipsychotic Drugs in First-Episode Schizophrenia and Schizophreniform Disorder: A Randomized, Open-Label Clinical Trial (EUFEST). *Am. J. Psychiatry* 166, 675–682. doi:10.1176/appi.ajp.2008.08060806
- Davis, J., Eyre, H., Jacka, F.N., Dodd, S., Dean, O., McEwen, S., Debnath, M., McGrath, J., Maes, M., Amminger, P., McGorry, P.D., Pantelis, C., Berk, M., 2016. A review of vulnerability and risks for schizophrenia: Beyond the two hit hypothesis. *Neurosci. Biobehav. Rev.* 65, 185–94.
doi:10.1016/j.neubiorev.2016.03.017
- Day, J., Fibiger, H.C., 1993. DOPAMINERGIC REGULATION OF CORTICAL ACETYLCHOLINE RELEASE: EFFECTS OF DOPAMINE RECEPTOR AGONISTS. *Neuroscience* 54, 643–648.
- De Angelis, C., Drazen, J.M., Frizelle, F.A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A.J.P.M., Schroeder, T. V., Sox, H.C., Van Der Weyden, M.B., 2004. Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors. *Ann. Intern. Med.* 141, 477.
doi:10.7326/0003-4819-141-6-200409210-00109
- de Vries, R.B.M., Wever, K.E., Avey, M.T., Stephens, M.L., Sena, E.S., Leenaars, M., 2014. The usefulness of systematic reviews of animal experiments for the design of preclinical and clinical studies. *ILAR J.* 55, 427–37.
doi:10.1093/ilar/ilu043
- Dean, D.J., Orr, J.M., Bernard, J.A., Gupta, T., Pelletier-Baldelli, A., Carol, E.E., Mittal, V.A., 2016. Hippocampal Shape Abnormalities Predict Symptom Progression in Neuroleptic-Free Youth at Ultrahigh Risk for Psychosis. *Schizophr. Bull.* 42, 161–9. doi:10.1093/schbul/sbv086
- Dean, K., Murray, R.M., 2005. Environmental risk factors for psychosis. *Dialogues Clin. Neurosci.* 7, 69–80.
- Debelle, M., Németh, G., Szalai, E., Szatmári, B., Harsányi, J., Barabassy, A., Laszlovszky, I., 2015. P.3.d.053 Cariprazine as monotherapy for the treatment of schizophrenia patients with predominant negative symptoms: a double-blind, active controlled trial. *Eur. Neuropsychopharmacol.* 25, S510.
doi:10.1016/S0924-977X(15)30701-X
- DerSimonian, R., Laird, N., 1986. Meta-analysis in clinical trials. *Control. Clin. Trials*

7, 177–88.

- Dietsche, B., Kircher, T., Falkenberg, I., 2017. Structural brain changes in schizophrenia at different stages of the illness: A selective review of longitudinal magnetic resonance imaging studies. *Aust. New Zeal. J. Psychiatry* 51, 500–508. doi:10.1177/0004867417699473
- DiMasi, J.A., Hansen, R.W., Grabowski, H.G., 2003. The price of innovation: new estimates of drug development costs. *J. Health Econ.* 22, 151–85. doi:10.1016/S0167-6296(02)00126-1
- Doherty, J.L., Owen, M.J., 2014. Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. *Genome Med.* 6, 29. doi:10.1186/gm546
- Donaldson, Z.R., Hen, R., 2015. From psychiatric disorders to animal models: a bidirectional and dimensional approach. *Biol. Psychiatry* 77, 15–21. doi:10.1016/j.biopsych.2014.02.004
- Duman, R.S., Li, N., Liu, R.-J., Duric, V., Aghajanian, G., 2012. Signaling pathways underlying the rapid antidepressant actions of ketamine. *Neuropharmacology* 62, 35–41. doi:10.1016/j.neuropharm.2011.08.044
- Dwan, K., Altman, D.G., Arnaiz, J.A., Bloom, J., Chan, A.-W., Cronin, E., Decullier, E., Easterbrook, P.J., Von Elm, E., Gamble, C., Gherzi, D., Ioannidis, J.P.A., Simes, J., Williamson, P.R., 2008. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. *PLoS One* 3, e3081. doi:10.1371/journal.pone.0003081
- Eagle, D.M., Robbins, T.W., 2003. Inhibitory Control in Rats Performing a Stop-Signal Reaction-Time Task: Effects of Lesions of the Medial Striatum and d-Amphetamine. *Behav. Neurosci.* 117, 1302–1317. doi:10.1037/0735-7044.117.6.1302
- Easterbrook, P.J., Berlin, J.A., Gopalan, R., Matthews, D.R., 1991. Publication bias in clinical research. *Lancet (London, England)* 337, 867–72.
- Egan, K.J., Vesterinen, H.M., Beglopoulos, V., Sena, E.S., Macleod, M.R., 2016. From a mouse: systematic analysis reveals limitations of experiments testing interventions in Alzheimer's disease mouse models. *Evidence-based Preclin. Med.* 3, e00015. doi:10.1002/ebm2.15
- Egan, M., Maclean, A., Sweeting, H., Hunt, K., 2012. Comparing the effectiveness of using generic and specific search terms in electronic databases to identify health outcomes for a systematic review: a prospective comparative study of literature search methods. *BMJ Open* 2, e001043. doi:10.1136/bmjopen-2012-001043
- Egger, M., Davey Smith, G., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315, 629–34.
- Eisener-Dorman, A.F., Grabowski-Boase, L., Tarantino, L.M., 2011. Cocaine locomotor activation, sensitization and place preference in six inbred strains of mice. *Behav. Brain Funct.* 7, 29. doi:10.1186/1744-9081-7-29

- Essali, A., Al-Haj Haasan, N., Li, C., Rathbone, J., 2009. Clozapine versus typical neuroleptic medication for schizophrenia. *Cochrane Database Syst. Rev.* doi:10.1002/14651858.CD000059.pub2
- European Monitoring Centre for Drugs and Drug Addiction, 2010. Guidelines for collecting data on retail drug prices in Europe : issues and challenges. Publications Office of the European Union, Luxembourg.
- Farrell, M.S., Werge, T., Sklar, P., Owen, M.J., Ophoff, R.A., O'Donovan, M.C., Corvin, A., Cichon, S., Sullivan, P.F., 2015. Evaluating historical candidate genes for schizophrenia. *Mol. Psychiatry* 20, 555–562. doi:10.1038/mp.2015.16
- Fatemi, S.H., Folsom, T.D., 2009. The neurodevelopmental hypothesis of schizophrenia, revisited. *Schizophr. Bull.* 35, 528–48. doi:10.1093/schbul/sbn187
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From Data Mining to Knowledge Discovery in Databases. *AI Mag.* 17, 37.
- Featherstone, R.E., Kapur, S., Fletcher, P.J., 2007. The amphetamine-induced sensitized state as a model of schizophrenia. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* 31, 1556–1571. doi:10.1016/J.PNPBP.2007.08.025
- Feifel, D., Shilling, P.D., 2010. Promise and pitfalls of animal models of schizophrenia. *Curr. Psychiatry Rep.* 12, 327–34. doi:10.1007/s11920-010-0122-x
- Fellner, C., 2017. New Schizophrenia Treatments Address Unmet Clinical Needs. *P T* 42, 130–134.
- Fernando, A.B.P., Robbins, T.W., 2011. Animal models of neuropsychiatric disorders. *Annu. Rev. Clin. Psychol.* 7, 39–61. doi:10.1146/annurev-clinpsy-032210-104454
- Festing, M.F.W., 2013. We are not born knowing how to design and analyse scientific experiments. *Altern. Lab. Anim.* 41, P19-21.
- Fisher, R.A., 1937. *The design of experiments*. New York : Hafner Press.
- Flinkkilä, E., Keski-Rahkonen, A., Marttunen, M., Raevuori, A., 2016. Prenatal Inflammation, Infections and Mental Disorders. *Psychopathology* 49, 317–333. doi:10.1159/000448054
- Frankel, J.S., Schwartz, T.L., 2017. Brexpiprazole and cariprazine: distinguishing two new atypical antipsychotics from the original dopamine stabilizer aripiprazole. *Ther. Adv. Psychopharmacol.* 7, 29–41. doi:10.1177/2045125316672136
- Freedman, R., Ross, R., Leonard, S., Myles-Worsley, M., Adams, C.E., Waldo, M., Tregellas, J., Martin, L., Olincy, A., Tanabe, J., Kisley, M.A., Hunter, S., Stevens, K.E., 2005. Early biomarkers of psychosis. *Dialogues Clin. Neurosci.* 7, 17–29.
- Friston, K., Brown, H.R., Siemerikus, J., Stephan, K.E., 2016. The dysconnection hypothesis (2016). *Schizophr. Res.* 176, 83–94.

- Garey, L.J., Ong, W.Y., Patel, T.S., Kanani, M., Davis, A., Mortimer, A.M., Barnes, T.R., Hirsch, S.R., 1998. Reduced dendritic spine density on cerebral cortical pyramidal neurons in schizophrenia. *J. Neurol. Neurosurg. Psychiatry* 65, 446–53.
- Garriga, M., Pacchiarotti, I., Kasper, S., Zeller, S.L., Allen, M.H., Vázquez, G., Baldaçara, L., San, L., McAllister-Williams, R.H., Fountoulakis, K.N., Courtet, P., Naber, D., Chan, E.W., Fagiolini, A., Möller, H.J., Grunze, H., Llorca, P.M., Jaffe, R.L., Yatham, L.N., Hidalgo-Mazzei, D., Passamar, M., Messer, T., Bernardo, M., Vieta, E., 2016. Assessment and management of agitation in psychiatry: Expert consensus. *World J. Biol. Psychiatry* 17, 86–128. doi:10.3109/15622975.2015.1132007
- Gaspar, P.A., Bosman, C., Ruiz, S., Aboitiz, F., 2009. The Aberrant Connectivity Hypothesis in Schizophrenia, in: Aboitiz, F., Cosmelli, D. (Eds.), *From Attention to Goal-Directed Behavior : Neurodynamical, Methodological, Clinical Trends*. Springer, Berlin.
- Gaur, N., Gautam, S., Gaur, M., Sharma, P., Dadheech, G., Mishra, S., 2008. The biochemical womb of schizophrenia: A review. *Indian J. Clin. Biochem.* 23, 307–327. doi:10.1007/s12291-008-0071-x
- Geerts, H., 2016. Drugs in Trials [WWW Document]. *Schizophr. Res. Forum*. URL <https://www.schizophreniaforum.org/drugs-trials> (accessed 2.23.18).
- Gerdin, A.-K., Igosheva, N., Roberson, L.-A., Ismail, O., Karp, N., Sanderson, M., Cambridge, E., Shannon, C., Sunter, D., Ramirez-Solis, R., Bussell, J., White, J.K., 2012. Experimental and husbandry procedures as potential modifiers of the results of phenotyping tests. *Physiol. Behav.* 106, 602–611. doi:10.1016/j.physbeh.2012.03.026
- Geyer, M.A., 2008. Developing translational animal models for symptoms of schizophrenia or bipolar mania. *Neurotox. Res.* 14, 71–78. doi:10.1007/BF03033576
- Geyer, M.A., 2006. Are cross-species measures of sensorimotor gating useful for the discovery of procognitive cotreatments for schizophrenia? *Dialogues Clin. Neurosci.* 8, 9–16.
- Geyer, M.A., Krebs-Thomson, K., Braff, D.L., Swerdlow, N.R., 2001. Pharmacological studies of prepulse inhibition models of sensorimotor gating deficits in schizophrenia: a decade in review. *Psychopharmacology (Berl.)* 156, 117–54.
- Geyer, M.A., Markou, A., 2000. Animal Models of Psychiatric Disorders, in: Bunney, B.S., Bloom, F.E. (Eds.), *Psychopharmacology : The Fourth Generation of Progress*. Raven Press, New York.
- Geyer, M.A., Moghaddam, B., 2002. Animal Models Relevant To Schizophrenia Disorders, in: Davis, K.L., American College of Neuropsychopharmacology. (Eds.), *Neuropsychopharmacology : the fifth generation of progress*. Lippincott Williams and Wilkins, Philadelphia [etc].

- Geyer, M.A., Olivier, B., Joëls, M., Kahn, R.S., 2012. From antipsychotic to anti-schizophrenia drugs: role of animal models. *Trends Pharmacol. Sci.* 33, 515–21. doi:10.1016/j.tips.2012.06.006
- Gibson, S., Brand, S.L., Burt, S., Boden, Z.V.R., Benson, O., 2013. Understanding treatment non-adherence in schizophrenia and bipolar disorder: a survey of what service users do and why. *BMC Psychiatry* 13, 153. doi:10.1186/1471-244X-13-153
- Glahn, D.C., Ragland, J.D., Abramoff, A., Barrett, J., Laird, A.R., Bearden, C.E., Velligan, D.I., 2005. Beyond hypofrontality: A quantitative meta-analysis of functional neuroimaging studies of working memory in schizophrenia. *Hum. Brain Mapp.* 25, 60–69. doi:10.1002/hbm.20138
- Gluud, L.L., 2006. Bias in Clinical Intervention Research. *Am. J. Epidemiol.* 163, 493–501. doi:10.1093/aje/kwj069
- Gobira, P.H., Ropke, J., Aguiar, D.C., Crippa, J.A.S., Moreira, F.A., 2013. Animal models for predicting the efficacy and side effects of antipsychotic drugs. *Rev. Bras. Psiquiatr.* 35, S132–S139. doi:10.1590/1516-4446-2013-1164
- Goetghebeur, P.J., Swartz, J.E., 2016. True alignment of preclinical and clinical research to enhance success in CNS drug development: a review of the current evidence. *J. Psychopharmacol.* 30, 586–594. doi:10.1177/0269881116645269
- Goff, D.C., Coyle, J.T., 2001. The Emerging Role of Glutamate in the Pathophysiology and Treatment of Schizophrenia. *Am. J. Psychiatry* 158, 1367–1377. doi:10.1176/appi.ajp.158.9.1367
- Goff, D.C., Henderson, D.C., Amico, E., 1992. Cigarette smoking in schizophrenia: relationship to psychopathology and medication side effects. *Am. J. Psychiatry* 149, 1189–1194. doi:10.1176/ajp.149.9.1189
- Grace, A.A., 2012. Dopamine system dysregulation by the hippocampus: implications for the pathophysiology and treatment of schizophrenia. *Neuropharmacology* 62, 1342–8. doi:10.1016/j.neuropharm.2011.05.011
- Grady, S.E., Marsh, T.A., Tenhouse, A., Klein, K., 2017. Ketamine for the treatment of major depressive disorder and bipolar depression: A review of the literature. *Ment. Heal. Clin.* 7, 16–23. doi:10.9740/mhc.2017.01.016
- Grant, M.J., 2013. What makes a good title? *Heal. Inf. Libr. J.* 30, 259–260. doi:10.1111/hir.12049
- Great Britain Department of Health, 2007. *Mental Health Act 2007 : consultation on secondary legislation.* Department of Health, London.
- Great Britain Office for National Statistics, 2002. *Psychiatric morbidity among adults living in private households, 2000 : summary report.*
- Green, M.F., Kern, R.S., Heaton, R.K., 2004. Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS. *Schizophr. Res.* 72, 41–51. doi:10.1016/j.schres.2004.09.009
- Greenwood, T.A., Light, G.A., Swerdlow, N.R., Calkins, M.E., Green, M.F., Gur,

- R.E., Gur, R.C., Lazzeroni, L.C., Nuechterlein, K.H., Olincy, A., Radant, A.D., Seidman, L.J., Siever, L.J., Silverman, J.M., Stone, W.S., Sugar, C.A., Tsuang, D.W., Tsuang, M.T., Turetsky, B.I., Freedman, R., Braff, D.L., 2016. Gating Deficit Heritability and Correlation With Increased Clinical Severity in Schizophrenia Patients With Positive Family History. *Am. J. Psychiatry* 173, 385–391. doi:10.1176/appi.ajp.2015.15050605
- Grossman, M.R., Cormack, G. V, 2011. Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review. *Richmond J. Law Technol.* XVII, 1–33. doi:10.1525/sp.2007.54.1.23.
- Gründer, G., Heinze, M., Cordes, J., Mühlbauer, B., Juckel, G., Schulz, C., Rüter, E., Timm, J., NeSSy Study Group, 2016. Effects of first-generation antipsychotics versus second-generation antipsychotics on quality of life in schizophrenia: a double-blind, randomised study. *The lancet. Psychiatry* 3, 717–729. doi:10.1016/S2215-0366(16)00085-7
- Gulin, J.E.N., Rocco, D.M., García-Bournissen, F., 2015. Quality of Reporting and Adherence to ARRIVE Guidelines in Animal Studies for Chagas Disease Preclinical Drug Research: A Systematic Review. *PLoS Negl. Trop. Dis.* 9, e0004194. doi:10.1371/journal.pntd.0004194
- Guloksuz, S., van Os, J., 2018. The slow death of the concept of schizophrenia and the painful birth of the psychosis spectrum. *Psychol. Med.* 48, 229–244. doi:10.1017/S0033291717001775
- Gupta, S., 2010. First-generation vs second-generation antipsychotic drugs: The ongoing saga. *Indian J. Psychiatry* 52, 77. doi:10.4103/0019-5545.58904
- Gurillo, P., Jauhar, S., Murray, R.M., MacCabe, J.H., 2015. Does tobacco use cause psychosis? Systematic review and meta-analysis. *The lancet. Psychiatry* 2, 718–725. doi:10.1016/S2215-0366(15)00152-2
- Hackam, D.G., Redelmeier, D.A., 2006. Translation of Research Evidence From Animals to Humans. *JAMA* 296, 1727. doi:10.1001/jama.296.14.1731
- Hadar, R., Soto-Montenegro, M.L., Götz, T., Wieske, F., Sohr, R., Desco, M., Hamani, C., Weiner, I., Pascau, J., Winter, C., 2015. Using a maternal immune stimulation model of schizophrenia to study behavioral and neurobiological alterations over the developmental course. *Schizophr. Res.* 166, 238–47. doi:10.1016/j.schres.2015.05.010
- Häfner, H., 2003. Gender differences in schizophrenia. *Psychoneuroendocrinology* 28 Suppl 2, 17–54.
- Halberstadt, A.L., Geyer, M.A., 2010. LSD but not lisuride disrupts prepulse inhibition in rats by activating the 5-HT(2A) receptor. *Psychopharmacology (Berl)*. 208, 179–89. doi:10.1007/s00213-009-1718-x
- Halladay, C.W., Trikalinos, T.A., Schmid, I.T., Schmid, C.H., Dahabreh, I.J., 2015. Using data sources beyond PubMed has a modest impact on the results of systematic reviews of therapeutic interventions. *J. Clin. Epidemiol.* 68, 1076–1084. doi:10.1016/j.jclinepi.2014.12.017
- Hånell, A., Marklund, N., 2014. Structured evaluation of rodent behavioral tests used

in drug discovery research. *Front. Behav. Neurosci.* 8, 252.
doi:10.3389/fnbeh.2014.00252

- Harris, E.C., Barraclough, B., 1998. Excess mortality of mental disorder. *Br. J. Psychiatry* 173, 11–53.
- Harrison, P.J., Weinberger, D.R., 2005. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol. Psychiatry* 10, 40–68.
doi:10.1038/sj.mp.4001558
- Hart, B., Lundh, A., Bero, L., 2012. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. *BMJ* 344, d7202–d7202.
doi:10.1136/bmj.d7202
- Hartling, L., Featherstone, R., Nuspl, M., Shave, K., Dryden, D.M., Vandermeer, B., 2016. The contribution of databases to the results of systematic reviews: a cross-sectional study. *BMC Med. Res. Methodol.* 16, 127. doi:10.1186/s12874-016-0232-1
- Harvey, P.D., 2009. Pharmacological Cognitive Enhancement in Schizophrenia. *Neuropsychol. Rev.* 19, 324–335. doi:10.1007/s11065-009-9103-4
- Harvey, P.D., Bowie, C.R., 2012. Cognitive enhancement in schizophrenia: pharmacological and cognitive remediation approaches. *Psychiatr. Clin. North Am.* 35, 683–98. doi:10.1016/j.psc.2012.06.008
- Hay, M., Thomas, D.W., Craighead, J.L., Economides, C., Rosenthal, J., 2014. Clinical development success rates for investigational drugs. *Nat. Biotechnol.* 32, 40–51. doi:10.1038/nbt.2786
- Hayes, E., Gavrilidis, E., Kulkarni, J., 2012. The role of oestrogen and other hormones in the pathophysiology and treatment of schizophrenia. *Schizophr. Res. Treatment* 2012, 540273. doi:10.1155/2012/540273
- Hearst, M.A., 1999. Untangling text data mining. *Proc. 37th Annu. Meet. Assoc. Comput. Linguist. Comput. Linguist.* 3–10. doi:10.3115/1034678.1034679
- Heckers, S., 2001. Neuroimaging studies of the hippocampus in schizophrenia. *Hippocampus* 11, 520–528. doi:10.1002/hipo.1068
- Hengartner, M.P., Lehmann, S.N., 2017. Why psychiatric research must abandon traditional diagnostic classification and adopt a fully dimensional scope: Two solutions to a persistent problem. *Front. Psychiatry* 8, 101.
doi:10.3389/fpsy.2017.00101
- Higgins, J.P.T., Altman, D.G., Gøtzsche, P.C., Jüni, P., Moher, D., Oxman, A.D., Savović, J., Schulz, K.F., Weeks, L., Sterne, J.A.C., 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343.
- Higgins, J.P.T., Green, S., 2008. *Cochrane handbook for systematic reviews of interventions, Systematic reviews of interventions.* Wiley-Blackwell.
- Hirst, T.C., Vesterinen, H.M., Conlin, S., Egan, K.J., Antonic, A., Lawson McLean, A., Macleod, M.R., Grant, R., Brennan, P.M., Sena, E.S., Whittle, I.R., 2014. A systematic review and meta-analysis of gene therapy in animal models of cerebral glioma: why did promise not translate to human therapy? *Evidence-*

based *Preclin. Med.* 1, 21–33. doi:10.1002/ebm2.6

Hoffman, K.L., 2016. Modeling Neuropsychiatric Disorders in Laboratory Animals, Modeling Neuropsychiatric Disorders in Laboratory Animals. doi:10.1016/B978-0-08-100099-1.00002-9

Homayoun, H., Moghaddam, B., 2007. NMDA receptor hypofunction produces opposite effects on prefrontal cortex interneurons and pyramidal neurons. *J. Neurosci.* 27, 11496–500. doi:10.1523/JNEUROSCI.2213-07.2007

Hooijmans, C.R., Leenaars, M., Ritskes-Hoitinga, M., 2010a. A gold standard publication checklist to improve the quality of animal studies, to fully integrate the Three Rs, and to make systematic reviews more feasible. *Altern. Lab. Anim.* 38, 167–82.

Hooijmans, C.R., Ritskes-Hoitinga, M., 2013. Progress in Using Systematic Reviews of Animal Studies to Improve Translational Research. *PLoS Med.* 10, e1001482. doi:10.1371/journal.pmed.1001482

Hooijmans, C.R., Rovers, M.M., de Vries, R.B., Leenaars, M., Ritskes-Hoitinga, M., Langendam, M.W., Sandercock, P., Roberts, I., Hooijmans, C., Rovers, M., Vries, R. de, Leenaars, M., Ritskes-Hoitinga, M., Muhlhausler, B., Bloomfield, F., Gillman, M., Chalmers, T., Celano, P., Sacks, H., Smith, H., Macleod, M., Fisher, M., O'Collins, V., Sena, E., Dirnagl, U., Bath, P., Buchan, A., Worp, H. van der, Traystman, R., Minematsu, K., Donnan, G., Howells, D., Moher, D., Cook, D., Jadad, A., Tugwell, P., Moher, M., Jones, A., Pham, B., Klassen, T., Schulz, K., Chalmers, I., Hayes, R., Altman, D., Higgins, J., Altman, D., Gotzsche, P., Juni, P., Moher, D., Oxman, A., Savovic, J., Schulz, K., Weeks, L., Sterne, J., Hooijmans, C., Jong, P.P., Vries, R. de, Ritskes-Hoitinga, M., Krauth, D., Woodruff, T., Bero, L., Sena, E., Worp, H. van der, Howells, D., Macleod, M., Wever, K., Menting, T., Rovers, M., Vliet, J. van der, Rongen, G., Masereeuw, R., Ritskes-Hoitinga, M., Hooijmans, C., Warle, M., Thayer, K., Rooney, A., Boyles, A., Holmgren, S., Walker, V., Kissling, G., Whiting, P., Rutjes, A., Westwood, M., Mallett, S., Deeks, J., Reitsma, J., Leeflang, M., Sterne, J., Bossuyt, P., Whiting, P., Rutjes, A., Reitsma, J., Bossuyt, P., Kleijnen, J., Kilkeny, C., Parsons, N., Kadyszewski, E., Festing, M., Cuthill, I., Fry, D., Hutton, J., Altman, D., ACG, K.B., Zutphen, L. van, Perel, P., Roberts, I., Sena, E., Wheble, P., Briscoe, C., Sandercock, P., Macleod, M., Mignini, L., Jayaram, P., Khan, K., Roberts, I., Kwan, I., Evans, P., Haig, S., Faggion, C., Giannakopoulos, N., Listl, S., Hooijmans, C., Vries, R. de, Rovers, M., Gooszen, H., Ritskes-Hoitinga, M., Hooijmans, C., Leenaars, M., Ritskes-Hoitinga, M., Kilkeny, C., Browne, W., Cuthill, I., Emerson, M., Altman, D., Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K., Simel, D., Stroup, D., Moher, D., Liberati, A., Tetzlaff, J., Altman, D., Johnston, N., Nevalainen, T., Kemppinen, N., Meller, A., Bjork, E., Kohila, T., Nevalainen, T., Donnelly, H., Saibaba, P., Vanderschuren, L., Niesink, R., Spruijt, B., Vanree, J., Clough, G., Clough, G., Harri, M., Bruguerolle, B., Bruguerolle, B., Valli, M., Bouyard, L., Jadot, G., Bouyard, P., Marrino, P., Gavish, D., Shafrir, E., Eisenberg, S., Claassen, V., Claassen, V., 2014. SYRCLE's risk of bias tool for animal studies. *BMC Med. Res. Methodol.* 14, 43. doi:10.1186/1471-2288-14-43

Hooijmans, C.R., Tillema, A., Leenaars, M., Ritskes-Hoitinga, M., 2010b. Enhancing search efficiency by means of a search filter for finding all studies on animal

- experimentation in PubMed. *Lab. Anim.* 44, 170–5.
doi:10.1258/la.2010.009117
- Horn, J., de Haan, R.J., Vermeulen, M., Luiten, P.G.M., Limburg, M., 2001. Nimodipine in Animal Model Experiments of Focal Cerebral Ischemia. *Stroke* 32.
- Howes, O., McCutcheon, R., Stone, J., 2015. Glutamate and dopamine in schizophrenia: an update for the 21st century. *J. Psychopharmacol.* 29, 97–115. doi:10.1177/0269881114563634
- Howes, O.D., Kapur, S., 2009. The dopamine hypothesis of schizophrenia: version III--the final common pathway. *Schizophr. Bull.* 35, 549–62.
doi:10.1093/schbul/sbp006
- Hurford, I.M., Marder, S.R., Keefe, R.S.E., Reise, S.P., Bilder, R.M., 2011. A Brief Cognitive Assessment Tool for Schizophrenia: Construction of a Tool for Clinicians. *Schizophr. Bull.* 37, 538–545. doi:10.1093/schbul/sbp095
- Hyman, S.E., 2010. The Diagnosis of Mental Disorders: The Problem of Reification. *Annu. Rev. Clin. Psychol.* 6, 155–179.
doi:10.1146/annurev.clinpsy.3.022806.091532
- Ichikawa, J., Ishii, H., Bonaccorso, S., Fowler, W.L., O’Laughlin, I.A., Meltzer, H.Y., 2001. 5-HT_{2A} and D₂ receptor blockade increases cortical DA release via 5-HT_{1A} receptor activation: a possible mechanism of atypical antipsychotic-induced cortical dopamine release. *J. Neurochem.* 76, 1521–1531.
doi:10.1046/j.1471-4159.2001.00154.x
- Ioannidis, J.P.A., 2014. How to Make More Published Research True. *PLoS Med.* 11, e1001747. doi:10.1371/journal.pmed.1001747
- Ioannidis, J.P.A., 2012. Extrapolating from animals to humans. *Sci. Transl. Med.* 4, 151ps15. doi:10.1126/scitranslmed.3004631
- Ioannidis, J.P.A., 2005. Why Most Published Research Findings Are False. *PLoS Med.* 2, e124. doi:10.1371/journal.pmed.0020124
- Ioannidis, J.P.A., Greenland, S., Hlatky, M.A., Khoury, M.J., Macleod, M.R., Moher, D., Schulz, K.F., Tibshirani, R., 2014a. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet (London, England)* 383, 166–75. doi:10.1016/S0140-6736(13)62227-8
- Ioannidis, J.P.A., Munafò, M.R., Fusar-Poli, P., Nosek, B.A., David, S.P., 2014b. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. *Trends Cogn. Sci.* 18, 235–41.
doi:10.1016/j.tics.2014.02.010
- Jackson, S.J., Andrews, N., Ball, D., Bellantuono, I., Gray, J., Hachoumi, L., Holmes, A., Latcham, J., Petrie, A., Potter, P., Rice, A., Ritchie, A., Stewart, M., Strepka, C., Yeoman, M., Chapman, K., 2017. Does age matter? The impact of rodent age on study outcomes. *Lab. Anim.* 51, 160–169.
doi:10.1177/0023677216653984
- Jacobus, J., Tapert, S.F., 2014. Effects of cannabis on the adolescent brain. *Curr.*

Pharm. Des. 20, 2186–93.

- Jahshan, C., Heaton, R.K., Golshan, S., Cadenhead, K.S., 2010. Course of neurocognitive deficits in the prodrome and first episode of schizophrenia. *Neuropsychology* 24, 109–20. doi:10.1037/a0016791
- Janowsky, D.S., el-Yousel, M.K., Davis, J.M., Sekerke, H.J., 1973. Provocation of schizophrenic symptoms by intravenous administration of methylphenidate. *Arch. Gen. Psychiatry* 28, 185–91.
- Javitt, D.C., Zukin, S.R., 1991. Recent advances in the phencyclidine model of schizophrenia. *Am. J. Psychiatry* 148, 1301–1308. doi:10.1176/ajp.148.10.1301
- Jelicic Kadic, A., Vucic, K., Dosenovic, S., Sapunar, D., Puljak, L., 2016. Extracting data from figures with software was faster, with higher interrater reliability than manual extraction. *J. Clin. Epidemiol.* 74, 119–123. doi:10.1016/j.jclinepi.2016.01.002
- Jia, P., Han, G., Zhao, J., Lu, P., Zhao, Z., 2017. SZGR 2.0: a one-stop shop of schizophrenia candidate genes. *Nucleic Acids Res.* 45, D915–D924. doi:10.1093/nar/gkw902
- Jones, C.A., Watson, D.J.G., Fone, K.C.F., 2011. Animal models of schizophrenia. *Br. J. Pharmacol.* 164, 1162–94. doi:10.1111/j.1476-5381.2011.01386.x
- Jonnalagadda, S.R., Goyal, P., Huffman, M.D., 2015. Automating data extraction in systematic reviews: a systematic review. *Syst. Rev.* 4, 78. doi:10.1186/s13643-015-0066-7
- Jovanovic, T., Blanding, N.Q., Norrholm, S.D., Duncan, E., Bradley, B., Ressler, K.J., 2009. Childhood abuse is associated with increased startle reactivity in adulthood. *Depress. Anxiety* 26, 1018–26. doi:10.1002/da.20599
- Jüni, P., Altman, D.G., Egger, M., 2001. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 323, 42–6.
- Kahn, R.S., Sommer, I.E., Murray, R.M., Meyer-Lindenberg, A., Weinberger, D.R., Cannon, T.D., O'Donovan, M., Correll, C.U., Kane, J.M., van Os, J., Insel, T.R., 2015. Schizophrenia. *Nat. Rev. Dis. Prim.* 1, 15067. doi:10.1038/nrdp.2015.67
- Kannan, G., Sawa, A., Pletnikov, M. V., 2013. Mouse models of gene–environment interactions in schizophrenia. *Neurobiol. Dis.* 57, 5–11. doi:10.1016/j.nbd.2013.05.012
- Kapur, S., Mamo, D., 2003. Half a century of antipsychotics and still a central role for dopamine D2 receptors. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 27, 1081–90. doi:10.1016/j.pnpbp.2003.09.004
- Kapur, S., Phillips, A.G., Insel, T.R., 2012. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol. Psychiatry* 17, 1174–1179. doi:10.1038/mp.2012.105
- Kapur, S., VanderSpek, S.C., Brownlee, B.A., Nobrega, J.N., 2003. Antipsychotic dosing in preclinical models is often unrepresentative of the clinical condition: a suggested solution based on in vivo occupancy. *J. Pharmacol. Exp. Ther.* 305,

625–31. doi:10.1124/jpet.102.046987

- Kapur, S., Wadenberg, M.-L., Remington, G., 2000. Are Animal Studies of Antipsychotics Appropriately Dosed?: Lessons from the Bedside to the Bench. *Can. J. Psychiatry* 45, 241–246. doi:10.1177/070674370004500302
- Karl, T., Arnold, J.C., 2014. Schizophrenia: a consequence of gene-environment interactions? *Front. Behav. Neurosci.* 8, 435. doi:10.3389/fnbeh.2014.00435
- Karlsgodt, K.H., Sun, D., Cannon, T.D., 2010. Structural and Functional Brain Abnormalities in Schizophrenia. *Curr. Dir. Psychol. Sci.* 19, 226–231. doi:10.1177/0963721410377601
- Karow, A., Wittmann, L., Schöttle, D., Schäfer, I., Lambert, M., 2014. The assessment of quality of life in clinical practice in patients with schizophrenia. *Dialogues Clin. Neurosci.* 16, 185–95.
- Kegeles, L.S., Abi-Dargham, A., Zea-Ponce, Y., Rodenhiser-Hill, J., Mann, J.J., Van Heertum, R.L., Cooper, T.B., Carlsson, A., Laruelle, M., 2000. Modulation of amphetamine-induced striatal dopamine release by ketamine in humans: implications for schizophrenia. *Biol. Psychiatry* 48, 627–40.
- Kelly, D.L., 2006. Treatment Considerations in Women with Schizophrenia. *J. Women's Heal.* 15, 1132–1140. doi:10.1089/jwh.2006.15.1132
- Kesby, J., Eyles, D., McGrath, J., Scott, J., 2018. Dopamine, psychosis and schizophrenia: the widening gap between basic and clinical neuroscience. *Transl. Psychiatry* 8, 30. doi:10.1038/s41398-017-0071-9
- Khandaker, G.M., Zimbron, J., Lewis, G., Jones, P.B., 2013. Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. *Psychol. Med.* 43, 239–57. doi:10.1017/S0033291712000736
- Khanna, P., Komossa, K., Rummel-Kluge, C., Hunger, H., Schwarz, S., El-Sayeh, H.G., Leucht, S., 2013. Aripiprazole versus other atypical antipsychotics for schizophrenia, in: Khanna, P. (Ed.), *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd, Chichester, UK, p. CD006569. doi:10.1002/14651858.CD006569.pub4
- Kilkenny, C., Browne, W., Cuthill, I.C., Emerson, M., Altman, D.G., NC3Rs Reporting Guidelines Working Group, 2010a. Animal research: reporting in vivo experiments: the ARRIVE guidelines. *Br. J. Pharmacol.* 160, 1577–9. doi:10.1111/j.1476-5381.2010.00872.x
- Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M., Altman, D.G., 2010b. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* 8, e1000412. doi:10.1371/journal.pbio.1000412
- Kilkenny, C., Parsons, N., Kadyszewski, E., Festing, M.F.W., Cuthill, I.C., Fry, D., Hutton, J., Altman, D.G., 2009. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One* 4, e7824. doi:10.1371/journal.pone.0007824
- Kirkbride, J.B., Errazuriz, A., Croudace, T.J., Morgan, C., Jackson, D., Boydell, J.,

- Murray, R.M., Jones, P.B., 2012. Incidence of Schizophrenia and Other Psychoses in England, 1950–2009: A Systematic Review and Meta-Analyses. *PLoS One* 7, e31660. doi:10.1371/journal.pone.0031660
- Kleinhaus, K., Harlap, S., Perrin, M., Manor, O., Weiser, M., Lichtenberg, P., Malaspina, D., 2011. Age, sex and first treatment of schizophrenia in a population cohort. *J. Psychiatr. Res.* 45, 136–41. doi:10.1016/j.jpsychires.2010.05.010
- Koenig, J., 2014. Animal Models for Schizophrenia Research [WWW Document]. *Schizophr. Res. Forum.* URL <https://www.schizophreniaforum.org/animal-models-schizophrenia-research> (accessed 2.23.18).
- Kokras, N., Dalla, C., 2014. Sex differences in animal models of psychiatric disorders. *Br. J. Pharmacol.* 171, 4595–619. doi:10.1111/bph.12710
- Kola, I., Landis, J., 2004. Opinion: Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3, 711–716. doi:10.1038/nrd1470
- Kolar, R., 2006. Animal experimentation. *Sci. Eng. Ethics* 12, 111–22.
- Konradi, C., Yang, C.K., Zimmerman, E.I., Lohmann, K.M., Gresch, P., Pantazopoulos, H., Berretta, S., Heckers, S., 2011. Hippocampal interneurons are abnormal in schizophrenia. *Schizophr. Res.* 131, 165–73. doi:10.1016/j.schres.2011.06.007
- Koukouli, F., Rooy, M., Tziotis, D., Sailor, K.A., O'Neill, H.C., Levenga, J., Witte, M., Nilges, M., Changeux, J.-P., Hoeffler, C.A., Stitzel, J.A., Gutkin, B.S., DiGregorio, D.A., Maskos, U., 2017. Nicotine reverses hypofrontality in animal models of addiction and schizophrenia. *Nat. Med.* 23, 347–354. doi:10.1038/nm.4274
- Krauth, D., Woodruff, T.J., Bero, L., 2013. Instruments for Assessing Risk of Bias and Other Methodological Criteria of Published Animal Studies: A Systematic Review. *Environ. Health Perspect.* doi:10.1289/ehp.1206389
- Krauthammer, M., Nenadic, G., 2004. Term identification in the biomedical literature. *J. Biomed. Inform.* 37, 512–526. doi:10.1016/j.jbi.2004.08.004
- Kubicki, M., McCarley, R.W., Shenton, M.E., 2005. Evidence for white matter abnormalities in schizophrenia. *Curr. Opin. Psychiatry* 18, 121–34.
- Kumar, M., Tripathi, C.D., Verma, V., Padhy, B.M., Meshram, G.G., Abhilash, B., 2016. Predictive Validity of Some Common Animal Models of Bipolar Disorder Using Lithium and Lamotrigine Therapy: An Attempt towards a Battery-Based Approach for the Evaluation of Mood Stabilizers. *Psychiatry Investig.* 13, 434–9. doi:10.4306/pi.2016.13.4.434
- Kuperberg, G.R., Broome, M.R., McGuire, P.K., David, A.S., Eddy, M., Ozawa, F., Goff, D., West, W.C., Williams, S.C.R., van der Kouwe, A.J.W., Salat, D.H., Dale, A.M., Fischl, B., 2003. Regionally Localized Thinning of the Cerebral Cortex in Schizophrenia. *Arch. Gen. Psychiatry* 60, 878. doi:10.1001/archpsyc.60.9.878
- Lally, J., Gaughran, F., Timms, P., Curran, S.R., 2016. Treatment-resistant

schizophrenia: current insights on the pharmacogenomics of antipsychotics. *Pharmacogenomics. Pers. Med.* 9, 117–129. doi:10.2147/PGPM.S115741

Lally, J., MacCabe, J.H., 2015. Antipsychotic medication in schizophrenia: a review. *Br. Med. Bull.* 114, 169–179. doi:10.1093/bmb/ldv017

Landis, S.C., Amara, S.G., Asadullah, K., Austin, C.P., Blumenstein, R., Bradley, E.W., Crystal, R.G., Darnell, R.B., Ferrante, R.J., Fillit, H., Finkelstein, R., Fisher, M., Gendelman, H.E., Golub, R.M., Goudreau, J.L., Gross, R.A., Gubitza, A.K., Hesterlee, S.E., Howells, D.W., Huguenard, J., Kelner, K., Koroshetz, W., Krainc, D., Lazic, S.E., Levine, M.S., Macleod, M.R., McCall, J.M., Moxley, R.T., Narasimhan, K., Noble, L.J., Perrin, S., Porter, J.D., Steward, O., Unger, E., Utz, U., Silberberg, S.D., 2012. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* 490, 187–91. doi:10.1038/nature11556

Larson, M.K., Walker, E.F., Compton, M.T., 2010. Early signs, diagnosis and therapeutics of the prodromal phase of schizophrenia and related psychotic disorders. *Expert Rev. Neurother.* 10, 1347–59. doi:10.1586/ern.10.93

Laruelle, M., 2000. The role of endogenous sensitization in the pathophysiology of schizophrenia: implications from recent brain imaging studies. *Brain Res. Brain Res. Rev.* 31, 371–84.

Laviolette, S.R., 2007. Dopamine Modulation of Emotional Processing in Cortical and Subcortical Neural Circuits: Evidence for a Final Common Pathway in Schizophrenia? *Schizophr. Bull.* 33, 971–981. doi:10.1093/schbul/sbm048

Lener, M.S., Wong, E., Tang, C.Y., Byne, W., Goldstein, K.E., Blair, N.J., Haznedar, M.M., New, A.S., Chemerinski, E., Chu, K.-W., Rimsky, L.S., Siever, L.J., Koenigsberg, H.W., Hazlett, E.A., 2015. White matter abnormalities in schizophrenia and schizotypal personality disorder. *Schizophr. Bull.* 41, 300–10. doi:10.1093/schbul/sbu093

Lenroot, R.K., Gogtay, N., Greenstein, D.K., Wells, E.M., Wallace, G.L., Clasen, L.S., Blumenthal, J.D., Lerch, J., Zijdenbos, A.P., Evans, A.C., Thompson, P.M., Giedd, J.N., 2007. Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *Neuroimage* 36, 1065–1073. doi:10.1016/j.neuroimage.2007.03.053

Leucht, S., Cipriani, A., Spineli, L., Mavridis, D., Örey, D., Richter, F., Samara, M., Barbui, C., Engel, R.R., Geddes, J.R., Kissling, W., Stapf, M.P., Lässig, B., Salanti, G., Davis, J.M., 2013. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet* 382, 951–962. doi:10.1016/S0140-6736(13)60733-3

Leucht, S., Corves, C., Arbter, D., Engel, R.R., Li, C., Davis, J.M., 2009. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. *Lancet* 373, 31–41. doi:10.1016/S0140-6736(08)61764-X

Leucht, S., Leucht, C., Huhn, M., Chaimani, A., Mavridis, D., Helfer, B., Samara, M., Rabaioli, M., Bächer, S., Cipriani, A., Geddes, J.R., Salanti, G., Davis, J.M., 2017. Sixty Years of Placebo-Controlled Antipsychotic Drug Trials in Acute Schizophrenia: Systematic Review, Bayesian Meta-Analysis, and Meta-Regression of Efficacy Predictors. *Am. J. Psychiatry* 174, 927–942.

doi:10.1176/appi.ajp.2017.16121358

- Leumann, L., Feldon, J., Vollenweider, F.X., Ludewig, K., 2002. Effects of typical and atypical antipsychotics on prepulse inhibition and latent inhibition in chronic schizophrenia. *Biol. Psychiatry* 52, 729–39.
- Leung, A., Chue, P., 2000. Sex differences in schizophrenia, a review of the literature. *Acta Psychiatr. Scand. Suppl.* 401, 3–38.
- Lewis, D., Volk, D., Hashimoto, T., 2004. Selective alterations in prefrontal cortical GABA neurotransmission in schizophrenia: a novel target for the treatment of working memory dysfunction. *Psychopharmacology (Berl)*. 174, 143–50. doi:10.1007/s00213-003-1673-x
- Li, J.-H., Kasinather, V., Cheung, Y.-W., Zhou, W., Nurhidayat, A.W., Jarlais, D., Schottenfeld, R., 2011. To use or not to use: an update on licit and illicit ketamine use. *Subst. Abuse Rehabil.* 2, 11. doi:10.2147/SAR.S15458
- Li, P., Snyder, G.L., Vanover, K.E., 2016. Dopamine Targeting Drugs for the Treatment of Schizophrenia: Past, Present and Future. *Curr. Top. Med. Chem.* 16, 3385–3403. doi:10.2174/1568026616666160608084834
- Lieberman, J.A., Kane, J.M., Alvir, J., 1987. Provocative tests with psychostimulant drugs in schizophrenia. *Psychopharmacology (Berl)*. 91, 415–433. doi:10.1007/BF00216006
- Light, G.A., Swerdlow, N.R., 2014. Neurophysiological Biomarkers Informing the Clinical Neuroscience of Schizophrenia: Mismatch Negativity and Prepulse Inhibition of Startle, in: *Current Topics in Behavioral Neurosciences*. pp. 293–314. doi:10.1007/7854_2014_316
- Light, G.A., Swerdlow, N.R., Rissling, A.J., Radant, A., Sugar, C.A., Sprock, J., Pela, M., Geyer, M.A., Braff, D.L., 2012. Characterization of Neurophysiologic and Neurocognitive Biomarkers for Use in Genomic and Clinical Outcome Studies of Schizophrenia. *PLoS One* 7, e39434. doi:10.1371/journal.pone.0039434
- Lloyd, K., White, J., 2011. Democratizing clinical research. *Nature* 474, 277–8. doi:10.1038/474277a
- Lo, B., Field, M.J., (U.S.), I. of M., Committee on Conflict of Interest in Medical Research and Practice., E., 2009. *Conflict of interest in medical research, education, and practice*. National Academies Press, Washington, D.C.
- Lodge, D.J., Behrens, M.M., Grace, A.A., 2009. A loss of parvalbumin-containing interneurons is associated with diminished oscillatory activity in an animal model of schizophrenia. *J. Neurosci.* 29, 2344–54. doi:10.1523/JNEUROSCI.5419-08.2009
- Lundh, A., Lexchin, J., Mintzes, B., Schroll, J.B., Bero, L., 2017. Industry sponsorship and research outcome. *Cochrane Database Syst. Rev.* 2, MR000033. doi:10.1002/14651858.MR000033.pub3
- MacDonald, P.L., Gardner, R.C., 2000. Type I Error Rate Comparisons of Post Hoc Procedures for I j Chi-Square Tables. *Educ. Psychol. Meas.* 60, 735–754. doi:10.1177/00131640021970871

- Macêdo, D.S., Araújo, D.P., Sampaio, L.R.L., Vasconcelos, S.M.M., Sales, P.M.G., Sousa, F.C.F., Hallak, J.E., Crippa, J.A., Carvalho, A.F., 2012. Animal models of prenatal immune challenge and their contribution to the study of schizophrenia: a systematic review. *Brazilian J. Med. Biol. Res. = Rev. Bras. Pesqui. medicas e Biol.* 45, 179–86. doi:10.1590/S0100-879X2012007500031
- Mackeprang, T., Kristiansen, K.T., Glenthøj, B.Y., 2002. Effects of antipsychotics on prepulse inhibition of the startle response in drug-naïve schizophrenic patients. *Biol. Psychiatry* 52, 863–73.
- Macleod, M.R., Fisher, M., O'Collins, V., Sena, E.S., Dirnagl, U., Bath, P.M.W., Buchan, A., Buchan, A., van der Worp, H.B., Traystman, R.J., Minematsu, K., Donnan, G.A., Howells, D.W., 2009. Reprint: Good laboratory practice: preventing introduction of bias at the bench. *J. Cereb. Blood Flow Metab.* 29, 221–3. doi:10.1038/jcbfm.2008.101
- Macleod, M.R., Fisher, M., O'Collins, V., Sena, E.S., Dirnagl, U., Bath, P.M.W., Buchan, A., van der Worp, H.B., Traystman, R., Minematsu, K., Donnan, G.A., Howells, D.W., 2009. Good Laboratory Practice. *Stroke* 40.
- Macleod, M.R., Lawson McLean, A., Kyriakopoulou, A., Serghiou, S., de Wilde, A., Sherratt, N., Hirst, T., Hemblade, R., Babor, Z., Nunes-Fonseca, C., Potluru, A., Thomson, A., Baginskitaie, J., Egan, K., Vesterinen, H., Currie, G.L., Churilov, L., Howells, D.W., Sena, E.S., 2015. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLoS Biol.* 13, 1–12. doi:10.1371/journal.pbio.1002273
- Macleod, M.R., O'Collins, T., Howells, D.W., Donnan, G.A., 2004. Pooling of Animal Experimental Data Reveals Influence of Study Design and Publication Bias. *Stroke* 35.
- Macleod, M.R., The NPQIP Collaborative group, 2017. Findings of a retrospective, controlled cohort study of the impact of a change in Nature journals' editorial policy for life sciences research on the completeness of reporting study design and execution. *bioRxiv* 187245. doi:10.1101/187245
- Macleod, M.R., van der Worp, H.B., Sena, E.S., Howells, D.W., Dirnagl, U., Donnan, G.A., 2008. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke.* 39, 2824–9. doi:10.1161/STROKEAHA.108.515957
- Maier, C., Baron, R., Tölle, T.R., Binder, A., Birbaumer, N., Birklein, F., Gierthmühlen, J., Flor, H., Geber, C., Hüge, V., Krumova, E.K., Landwehrmeyer, G.B., Magerl, W., Maihöfner, C., Richter, H., Rolke, R., Scherens, A., Schwarz, A., Sommer, C., Tronnier, V., Uçeyler, N., Valet, M., Wasner, G., Treede, R.-D., 2010. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): somatosensory abnormalities in 1236 patients with different neuropathic pain syndromes. *Pain* 150, 439–50. doi:10.1016/j.pain.2010.05.002
- Mailman, R.B., Murthy, V., 2010. Third generation antipsychotic drugs: partial agonism or receptor functional selectivity? *Curr. Pharm. Des.* 16, 488–501.
- Malhotra, D., McCarthy, S., Michaelson, J.J., Vacic, V., Burdick, K.E., Yoon, S., Cichon, S., Corvin, A., Gary, S., Gershon, E.S., Gill, M., Karayiorgou, M.,

- Kelsoe, J.R., Krastoshevsky, O., Krause, V., Leibenluft, E., Levy, D.L., Makarov, V., Bhandari, A., Malhotra, A.K., McMahon, F.J., Nöthen, M.M., Potash, J.B., Rietschel, M., Schulze, T.G., Sebat, J., 2011. High Frequencies of De Novo CNVs in Bipolar Disorder and Schizophrenia. *Neuron* 72, 951–963. doi:10.1016/j.neuron.2011.11.007
- Marcotte, E.R., Pearson, D.M., Srivastava, L.K., 2001. Animal models of schizophrenia: a critical review. *J. Psychiatry Neurosci.* 26, 395–410.
- Marek, G., Merchant, K., 2005. Developing therapeutics for schizophrenia and other psychotic disorders. *NeuroRx* 2, 579–89. doi:10.1602/neurorx.2.4.579
- Markou, A., Chiamulera, C., Geyer, M.A., Tricklebank, M., Steckler, T., 2009. Removing obstacles in neuroscience drug discovery: the future path for animal models. *Neuropsychopharmacology* 34, 74–89. doi:10.1038/npp.2008.173
- Marshall, I.J., Kuiper, J., Wallace, B.C., 2015. Automating Risk of Bias Assessment for Clinical Trials. *IEEE J. Biomed. Heal. Informatics* 19, 1406–1412. doi:10.1109/JBHI.2015.2431314
- Marsman, A., Mandl, R.C.W., Klomp, D.W.J., Bohlken, M.M., Boer, V.O., Andreychenko, A., Cahn, W., Kahn, R.S., Luijten, P.R., Hulshoff Pol, H.E., 2014. GABA and glutamate in schizophrenia: a 7 T ¹H-MRS study. *NeuroImage. Clin.* 6, 398–407. doi:10.1016/j.nicl.2014.10.005
- Mattei, D., Schweibold, R., Wolf, S.A., 2015. Brain in flames - animal models of psychosis: utility and limitations. *Neuropsychiatr. Dis. Treat.* 11, 1313–29. doi:10.2147/NDT.S65564
- Mauri, M.C., Paletta, S., Maffini, M., Colasanti, A., Dragogna, F., Di Pace, C., Altamura, A.C., 2014. Clinical pharmacology of atypical antipsychotics: an update. *EXCLI J.* 13, 1163–91.
- Maynard, T.M., Sikich, L., Lieberman, J.A., LaMantia, A.S., 2001. Neural development, cell-cell signaling, and the “two-hit” hypothesis of schizophrenia. *Schizophr. Bull.* 27, 457–76.
- McCann, S.K., Sena, E.S., Currie, G.L., Macleod, M.R., Howells, D.W., 2017. Systematic Review and Meta-analysis: Important Tools in Understanding Drug Development for Stroke. Springer, Cham, pp. 73–93. doi:10.1007/978-3-319-45345-3_2
- McGorry, P., Bates, T., Birchwood, M., 2013. Designing youth mental health services for the 21st century: examples from Australia, Ireland and the UK. *Br. J. Psychiatry* 202, s30–s35. doi:10.1192/bjp.bp.112.119214
- McGorry, P.D., 2015. Early intervention in psychosis: obvious, effective, overdue. *J. Nerv. Ment. Dis.* 203, 310–8. doi:10.1097/NMD.0000000000000284
- McGrath, J., Castle, D., Murray, R., 1995. How Can We Judge Whether Or Not Prenatal Exposure to Influenza Causes Schizophrenia?, in: *Neural Development and Schizophrenia*. Springer US, Boston, MA, pp. 203–214. doi:10.1007/978-1-4615-1955-3_13
- McGrath, J.J., Saha, S., Al-Hamzawi, A., Andrade, L., Benjet, C., Bromet, E.J.,

- Browne, M.O., Caldas de Almeida, J.M., Chiu, W.T., Demyttenaere, K., Fayyad, J., Florescu, S., de Girolamo, G., Gureje, O., Haro, J.M., ten Have, M., Hu, C., Kovess-Masfety, V., Lim, C.C.W., Navarro-Mateu, F., Sampson, N., Posada-Villa, J., Kendler, K.S., Kessler, R.C., 2016. The Bidirectional Associations Between Psychotic Experiences and DSM-IV Mental Disorders. *Am. J. Psychiatry* 173, 997–1006. doi:10.1176/appi.ajp.2016.15101293
- Mchedlidze, O., Dzadzamia, S., Butskhrikidze, M., Tsomaia, V., Nachkebia, N., 2011. Changes of locomotor, exploratory and emotional behavior in animal model of depression induced by deficiency of brain monoamine content. *Georgian Med. News* 76–82.
- Meincke, U., Mörth, D., Voß, T., Thelen, B., Geyer, M.A., Gouzoulis-Mayfrank, E., 2004. Prepulse inhibition of the acoustically evoked startle reflex in patients with an acute schizophrenic psychosis—A longitudinal study. *Eur. Arch. Psychiatry Clin. Neurosci.* 254, 415–421. doi:10.1007/s00406-004-0523-0
- Mena, A., Ruiz-Salas, J.C., Puentes, A., Dorado, I., Ruiz-Veguilla, M., De la Casa, L.G., 2016. Reduced Prepulse Inhibition as a Biomarker of Schizophrenia. *Front. Behav. Neurosci.* 10, 202. doi:10.3389/fnbeh.2016.00202
- Meng, Z.H., Feldpaush, D.L., Merchant, K.M., 1998. Clozapine and haloperidol block the induction of behavioral sensitization to amphetamine and associated genomic responses in rats. *Brain Res. Mol. Brain Res.* 61, 39–50.
- Millan, M.J., Agid, Y., Brüne, M., Bullmore, E.T., Carter, C.S., Clayton, N.S., Connor, R., Davis, S., Deakin, B., DeRubeis, R.J., Dubois, B., Geyer, M.A., Goodwin, G.M., Gorwood, P., Jay, T.M., Joëls, M., Mansuy, I.M., Meyer-Lindenberg, A., Murphy, D., Rolls, E., Saletu, B., Spedding, M., Sweeney, J., Whittington, M., Young, L.J., 2012. Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nat. Rev. Drug Discov.* 11, 141–168. doi:10.1038/nrd3628
- Millard, L.A.C., Flach, P.A., Higgins, J.P.T., 2016. Machine learning to assist risk-of-bias assessments in systematic reviews. *Int. J. Epidemiol.* 45, 266–77. doi:10.1093/ije/dyv306
- Minassian, A., Feifel, D., Perry, W., 2007. The relationship between sensorimotor gating and clinical improvement in acutely ill schizophrenia patients. *Schizophr. Res.* 89, 225–31. doi:10.1016/j.schres.2006.08.006
- Mintz, J., Kopelowicz, A., 2007. CULASS confirms CATIE. *Arch. Gen. Psychiatry* 64, 978; author reply 979-80. doi:10.1001/archpsyc.64.8.978-a
- Miyamoto, S., Duncan, G.E., Marx, C.E., Lieberman, J.A., 2005. Treatments for schizophrenia: a critical review of pharmacology and mechanisms of action of antipsychotic drugs. *Mol. Psychiatry* 10, 79–104. doi:10.1038/sj.mp.4001556
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* 4, 1. doi:10.1186/2046-4053-4-1
- Mokhtari, M., Rajarethinam, R., 2013. Early intervention and the treatment of prodrome in schizophrenia: A review of recent developments. *J. Psychiatr.*

Pract. doi:10.1097/01.pra.0000435036.83426.94

- Mondelli, V., Pariante, C.M., Navari, S., Aas, M., D'Albenzio, A., Di Forti, M., Handley, R., Hepgul, N., Marques, T.R., Taylor, H., Papadopoulos, A.S., Aitchison, K.J., Murray, R.M., Dazzan, P., 2010. Higher cortisol levels are associated with smaller left hippocampal volume in first-episode psychosis. *Schizophr. Res.* 119, 75–78. doi:10.1016/j.schres.2009.12.021
- Moore, H., 2010. The role of rodent models in the discovery of new treatments for schizophrenia: updating our strategy. *Schizophr. Bull.* 36, 1066–72. doi:10.1093/schbul/sbq106
- Moore, H., West, A.R., Grace, A.A., 1999. The regulation of forebrain dopamine transmission: relevance to the pathophysiology and psychopathology of schizophrenia. *Biol. Psychiatry* 46, 40–55.
- Moore, T.H.M., Zammit, S., Lingford-Hughes, A., Barnes, T.R.E., Jones, P.B., Burke, M., Lewis, G., 2007. Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. *Lancet (London, England)* 370, 319–28. doi:10.1016/S0140-6736(07)61162-3
- Morgan, C., Charalambides, M., Hutchinson, G., Murray, R.M., 2010. Migration, Ethnicity, and Psychosis: Toward a Sociodevelopmental Model. *Schizophr. Bull.* 36, 655–664. doi:10.1093/schbul/sbq051
- Morgan, S., Grootendorst, P., Lexchin, J., Cunningham, C., Greyson, D., 2011. The cost of drug development: A systematic review. *Health Policy (New York)*. 100, 4–17. doi:10.1016/j.healthpol.2010.12.002
- Mouri, A., Nagai, T., Ibi, D., Yamada, K., 2013. Animal models of schizophrenia for molecular and pharmacological intervention and potential candidate molecules. *Neurobiol. Dis.* 53, 61–74. doi:10.1016/J.NBD.2012.10.025
- Mueller, K.F., Briel, M., Strech, D., Meerpohl, J.J., Lang, B., Motschall, E., Gloy, V., Lamontagne, F., Bassler, D., 2014. Dissemination bias in systematic reviews of animal research: a systematic review. *PLoS One* 9, e116016. doi:10.1371/journal.pone.0116016
- Munro, C.A., McCaul, M.E., Wong, D.F., Oswald, L.M., Zhou, Y., Brasic, J., Kuwabara, H., Kumar, A., Alexander, M., Ye, W., Wand, G.S., 2006. Sex Differences in Striatal Dopamine Release in Healthy Adults. *Biol. Psychiatry* 59, 966–974. doi:10.1016/j.biopsych.2006.01.008
- Murray, R.M., Lewis, S.W., 1987. Is schizophrenia a neurodevelopmental disorder? *Br. Med. J. (Clin. Res. Ed)*. 295, 681–2.
- Myers, M.G., Rohsenow, D.J., Monti, P.M., Dey, A., 1995. Patterns of cocaine use among individuals in substance abuse treatment. *Am. J. Drug Alcohol Abuse* 21, 223–31.
- National Collaborating Centre for Mental Health (Great Britain), 2014. *Psychosis and Schizophrenia in Adults : Treatment and Management*.
- Nestler, E.J., Hyman, S.E., 2010. Animal models of neuropsychiatric disorders. *Nat. Neurosci.* 13, 1161–9. doi:10.1038/nn.2647

- NHS National Services Scotland, 2017. Medicines used in Mental Health - Years 2006/07 to 2016/17.
- Noordzij, M., Hooft, L., Dekker, F.W., Zoccali, C., Jager, K.J., 2009. Systematic reviews and meta-analyses: when they are useful and when to be careful. *Kidney Int.* 76, 1130–1136. doi:10.1038/ki.2009.339
- Nuechterlein, K.H., Barch, D.M., Gold, J.M., Goldberg, T.E., Green, M.F., Heaton, R.K., 2004. Identification of separable cognitive factors in schizophrenia. *Schizophr. Res.* 72, 29–39. doi:10.1016/j.schres.2004.09.007
- Nutt, D.J., Need, A.C., 2014. Where now for schizophrenia research? *Eur. Neuropsychopharmacol.* 24, 1181–1187. doi:10.1016/J.EURONEURO.2014.05.012
- Ochoa, S., Usall, J., Cobo, J., Labad, X., Kulkarni, J., 2012. Gender Differences in Schizophrenia and First-Episode Psychosis: A Comprehensive Literature Review. *Schizophr. Res. Treatment* 2012, 1–9. doi:10.1155/2012/916198
- Onstad, S., Skre, I., Edvardsen, J., Torgersen, S., Kringlen, E., 1991. Mental disorders in first-degree relatives of schizophrenics. *Acta Psychiatr. Scand.* 83, 463–467. doi:10.1111/j.1600-0447.1991.tb05577.x
- Oranje, B., Van Oel, C.J., Gispen-De Wied, C.C., Verbaten, M.N., Kahn, R.S., 2002. Effects of typical and atypical antipsychotics on the prepulse inhibition of the startle reflex in patients with schizophrenia. *J. Clin. Psychopharmacol.* 22, 359–65.
- Orellana, G., Slachevsky, A., 2013. Executive functioning in schizophrenia. *Front. psychiatry* 4, 35. doi:10.3389/fpsy.2013.00035
- Owen, M.J., 2014. New Approaches to Psychiatric Diagnostic Classification. *Neuron* 84, 564–571. doi:10.1016/J.NEURON.2014.10.028
- Owen, M.J., Sawa, A., Mortensen, P.B., 2016. Schizophrenia. *Lancet* 388, 86–97. doi:10.1016/S0140-6736(15)01121-6
- Paulson, P.E., Camp, D.M., Robinson, T.E., 1991. Time course of transient behavioral depression and persistent behavioral sensitization in relation to regional brain monoamine concentrations during amphetamine withdrawal in rats. *Psychopharmacology (Berl)*. 103, 480–92.
- Pennybaker, S.J., Luckenbaugh, D.A., Park, L.T., Marquardt, C.A., Zarate, C.A., 2017. Ketamine and Psychosis History: Antidepressant Efficacy and Psychotomimetic Effects Postinfusion. *Biol. Psychiatry* 82, e35–e36. doi:10.1016/j.biopsych.2016.08.041
- Percie du Sert, N., Bamsey, I., Bate, S.T., Berdoy, M., Clark, R.A., Cuthill, I., Fry, D., Karp, N.A., Macleod, M., Moon, L., Stanford, S.C., Lings, B., 2017. The Experimental Design Assistant. *PLOS Biol.* 15, e2003779. doi:10.1371/journal.pbio.2003779
- Perriol, M.-P., Dujardin, K., Derambure, P., Marcq, A., Bourriez, J.-L., Laureau, E., Pasquier, F., Defebvre, L., Destée, A., 2005. Disturbance of sensory filtering in dementia with Lewy bodies: comparison with Parkinson's disease dementia

- and Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* 76, 106–8.
doi:10.1136/jnnp.2003.035022
- Perry, W., Minassian, A., Feifel, D., Braff, D.L., 2001. Sensorimotor gating deficits in bipolar disorder patients with acute psychotic mania. *Biol. Psychiatry* 50, 418–24.
- Perry, W., Minassian, A., Henry, B., Kincaid, M., Young, J.W., Geyer, M.A., 2010. Quantifying over-activity in bipolar and schizophrenia patients in a human open field paradigm. *Psychiatry Res.* 178, 84–91.
doi:10.1016/j.psychres.2010.04.032
- Pettersson-Yeo, W., Allen, P., Benetti, S., McGuire, P., Mechelli, A., 2011. Dysconnectivity in schizophrenia: Where are we now? *Neurosci. Biobehav. Rev.* 35, 1110–1124. doi:10.1016/j.neubiorev.2010.11.004
- Pilowsky, L.S., Bressan, R.A., Stone, J.M., Erlandsson, K., Mulligan, R.S., Krystal, J.H., Ell, P.J., 2006. First in vivo evidence of an NMDA receptor deficit in medication-free schizophrenic patients. *Mol. Psychiatry* 11, 118–119.
doi:10.1038/sj.mp.4001751
- Piontkewitz, Y., Arad, M., Weiner, I., 2012. Tracing the development of psychosis and its prevention: what can be learned from animal models. *Neuropharmacology* 62, 1273–89. doi:10.1016/j.neuropharm.2011.04.019
- Plint, A.C., Moher, D., Morrison, A., Schulz, K., Altman, D.G., Hill, C., Gaboury, I., 2006. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med. J. Aust.* 185, 263–7.
- Pogarell, O., Koch, W., Karch, S., Dehning, S., Müller, N., Tatsch, K., Poepperl, G., Möller, H.-J., 2012. Dopaminergic Neurotransmission in Patients with Schizophrenia in Relation to Positive and Negative Symptoms. *Pharmacopsychiatry* 45, S36–S41. doi:10.1055/s-0032-1306313
- Poole, T., 1997. Happy animals make good science. *Lab. Anim.* 31, 116–124.
doi:10.1258/002367797780600198
- Post, R.M., 1975. Cocaine psychoses: a continuum model. *Am. J. Psychiatry* 132, 225–231. doi:10.1176/ajp.132.3.225
- Potvin, S., Stip, E., Roy, J., 2005. Toxic Psychoses as Pharmacological Models of Schizophrenia. *Curr. Psychiatry Rev.* 1, 23–32.
doi:10.2174/1573400052953583
- Powell, C.M., Miyakawa, T., 2006. Schizophrenia-relevant behavioral testing in rodent models: a uniquely human disorder? *Biol. Psychiatry* 59, 1198–207.
doi:10.1016/j.biopsych.2006.05.008
- Powell, S.B., 2010. Models of neurodevelopmental abnormalities in schizophrenia. *Curr. Top. Behav. Neurosci.* 4, 435–81.
- Prager, E.M., Bergstrom, H.C., Grunberg, N.E., Johnson, L.R., 2011. The importance of reporting housing and husbandry in rat research. *Front. Behav. Neurosci.* 5, 38. doi:10.3389/fnbeh.2011.00038
- Pratt, J., Winchester, C., Dawson, N., Morris, B., 2012. Advancing schizophrenia

drug discovery: optimizing rodent models to bridge the translational gap. *Nat. Rev. Drug Discov.* 11, 560–579. doi:10.1038/nrd3649

- Prescott, M.J., Lidster, K., 2017. Improving quality of science through better animal welfare: the NC3Rs strategy. *Nat. Publ. Gr.* 46. doi:10.1038/labam.1217
- Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E., Kähler, A., Duncan, L., Stahl, E., Genovese, G., Fernández, E., Collins, M.O., Komiyama, N.H., Choudhary, J.S., Magnusson, P.K.E., Banks, E., Shakir, K., Garimella, K., Fennell, T., Depristo, M., Grant, S.G.N., Haggarty, S.J., Gabriel, S., Scolnick, E.M., Lander, E.S., Hultman, C.M., Sullivan, P.F., McCarroll, S.A., Sklar, P., 2014. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 506, 185–190. doi:10.1038/nature12975
- Quednow, B.B., Frommann, I., Berning, J., Kühn, K.-U., Maier, W., Wagner, M., 2008. Impaired Sensorimotor Gating of the Acoustic Startle Response in the Prodrome of Schizophrenia. *Biol. Psychiatry* 64, 766–773. doi:10.1016/j.biopsych.2008.04.019
- Rai, S., Pathak, A., Sharma, I., 2015. Postpartum psychiatric disorders: Early diagnosis and management. *Indian J. Psychiatry* 57, S216-21. doi:10.4103/0019-5545.161481
- Rao, N.P., Kalmady, S., Arasappa, R., Venkatasubramanian, G., 2010. Clinical correlates of thalamus volume deficits in anti-psychotic-naïve schizophrenia patients: A 3-Tesla MRI study. *Indian J. Psychiatry* 52, 229–35. doi:10.4103/0019-5545.70975
- Ratajczak, P., Wozniak, A., Nowakowska, E., 2013. Animal models of schizophrenia: developmental preparation in rats. *Acta Neurobiol. Exp. (Wars)*. 73, 472–84.
- Ratnayake, U., Quinn, T.A., Castillo-Melendez, M., Dickinson, H., Walker, D.W., 2012. Behaviour and hippocampus-specific changes in spiny mouse neonates after treatment of the mother with the viral-mimetic Poly I:C at mid-pregnancy. *Brain. Behav. Immun.* 26, 1288–1299. doi:10.1016/j.bbi.2012.08.011
- Ravanic, D.B., Dejanovic, S.M.D., Janjic, V., Jovic, S.D., Milovanovic, D.R., Jakovljevic, V., Pantovic, V., Ravanic, B., Pantovic, M., Pantovic, M.M., 2009. Effectiveness of clozapine, haloperidol and chlorpromazine in schizophrenia during a five-year period. *Arq. Neuropsiquiatr.* 67, 195–202.
- Rees, E., O'Donovan, M.C., Owen, M.J., 2015. Genetics of schizophrenia. *Curr. Opin. Behav. Sci.* 2, 8–14. doi:10.1016/J.COBEHA.2014.07.001
- Reeves, R., Thiruchelvam, M., Cory-Slechta, D.A., 2004. Expression of behavioral sensitization to the cocaine-like fungicide triadimefon is blocked by pretreatment with AMPA, NMDA and DA D1 receptor antagonists. *Brain Res.* 1008, 155–167. doi:10.1016/j.brainres.2004.01.079
- Remington, G., Foussias, G., Fervaha, G., Agid, O., Takeuchi, H., Lee, J., Hahn, M., 2016. Treating Negative Symptoms in Schizophrenia: an Update. *Curr. Treat. options psychiatry* 3, 133–150. doi:10.1007/s40501-016-0075-8

- Remington, G., Kapur, S., 2010. Antipsychotic dosing: how much but also how often? *Schizophr. Bull.* 36, 900–3. doi:10.1093/schbul/sbq083
- Remington, G., Kapur, S., Foussias, G., Agid, O., Mann, S., Borlido, C., Richards, S., Javaid, N., 2012. Tetrabenazine Augmentation in Treatment-Resistant Schizophrenia. *J. Clin. Psychopharmacol.* 32, 95–99. doi:10.1097/JCP.0b013e31823f913e
- Richtand, N.M., 2006. Behavioral Sensitization, Alternative Splicing and D3 Dopamine Receptor-Mediated Inhibitory Function. *Neuropsychopharmacology* 31, 2368–2375. doi:10.1038/sj.npp.1301163
- Riecher-Rössler, A., 2002. Oestrogen effects in schizophrenia and their potential therapeutic implications - Review. *Arch. Womens. Ment. Health* 5, 111–118. doi:10.1007/s00737-002-0003-3
- Rietschel, L., Lambert, M., Karow, A., Zink, M., Müller, H., Heinz, A., de Millas, W., Janssen, B., Gaebel, W., Schneider, F., Naber, D., Juckel, G., Krüger-Özgürdal, S., Wobrock, T., Wagner, M., Maier, W., Klosterkötter, J., Bechdorf, A., PREVENT study group, 2017. Clinical high risk for psychosis: gender differences in symptoms and social functioning. *Early Interv. Psychiatry* 11, 306–313. doi:10.1111/eip.12240
- Ring, N., Tantam, D., Montague, L., Newby, D., Black, D., Morris, J., 1991. Gender differences in the incidence of definite schizophrenia and atypical psychosis? focus on negative symptoms of schizophrenia. *Acta Psychiatr. Scand.* 84, 489–496. doi:10.1111/j.1600-0447.1991.tb03182.x
- Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kähler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., Kim, Y., Lee, S.H., Magnusson, P.K.E., Sanchez, N., Stahl, E.A., Williams, S., Wray, N.R., Xia, K., Bettella, F., Borglum, A.D., Bulik-Sullivan, B.K., Cormican, P., Craddock, N., de Leeuw, C., Durmishi, N., Gill, M., Golimbet, V., Hamshere, M.L., Holmans, P., Hougaard, D.M., Kendler, K.S., Lin, K., Morris, D.W., Mors, O., Mortensen, P.B., Neale, B.M., O'Neill, F.A., Owen, M.J., Milovancevic, M.P., Posthuma, D., Powell, J., Richards, A.L., Riley, B.P., Ruderfer, D., Rujescu, D., Sigurdsson, E., Silagadze, T., Smit, A.B., Stefansson, H., Steinberg, S., Suvisaari, J., Tosato, S., Verhage, M., Walters, J.T., Levinson, D.F., Gejman, P. V., Kendler, K.S., Laurent, C., Mowry, B.J., O'Donovan, M.C., Owen, M.J., Pulver, A.E., Riley, B.P., Schwab, S.G., Wildenauer, D.B., Dudbridge, F., Holmans, P., Shi, J., Albus, M., Alexander, M., Campion, D., Cohen, D., Dikeos, D., Duan, J., Eichhammer, P., Godard, S., Hansen, M., Lerer, F.B., Liang, K.-Y., Maier, W., Mallet, J., Nertney, D.A., Nestadt, G., Norton, N., O'Neill, F.A., Papadimitriou, G.N., Ribble, R., Sanders, A.R., Silverman, J.M., Walsh, D., Williams, N.M., Wormley, B., Arranz, M.J., Bakker, S., Bender, S., Bramon, E., Collier, D., Crespo-Facorro, B., Hall, J., Iyegbe, C., Jablensky, A., Kahn, R.S., Kalaydjieva, L., Lawrie, S., Lewis, C.M., Lin, K., Linszen, D.H., Mata, I., McIntosh, A., Murray, R.M., Ophoff, R.A., Powell, J., Rujescu, D., Van Os, J., Walshe, M., Weisbrod, M., Wiersma, D., Donnelly, P., Barroso, I., Blackwell, J.M., Bramon, E., Brown, M.A., Casas, J.P., Corvin, A.P., Deloukas, P., Duncanson, A., Jankowski, J., Markus, H.S., Mathew, C.G., Palmer, C.N.A., Plomin, R., Rautanen, A., Sawcer, S.J., Trembath, R.C., Viswanathan, A.C., Wood, N.W., Spencer, C.C.A., Band, G., Bellenguez, C., Freeman, C., Hellenthal, G., Giannoulatou, E., Pirinen, M., Pearson, R.D., Strange, A., Su, Z., Vukcevic, D.,

- Donnelly, P., Langford, C., Hunt, S.E., Edkins, S., Gwilliam, R., Blackburn, H., Bumpstead, S.J., Dronov, S., Gillman, M., Gray, E., Hammond, N., Jayakumar, A., McCann, O.T., Liddle, J., Potter, S.C., Ravindrarajah, R., Ricketts, M., Tashakkori-Ghanbaria, A., Waller, M.J., Weston, P., Widaa, S., Whittaker, P., Barroso, I., Deloukas, P., Mathew, C.G., Blackwell, J.M., Brown, M.A., Corvin, A.P., McCarthy, M.I., Spencer, C.C.A., Bramon, E., Corvin, A.P., O'Donovan, M.C., Stefansson, K., Scolnick, E., Purcell, S., McCarroll, S.A., Sklar, P., Hultman, C.M., Sullivan, P.F., Sklar, P., Hultman, C.M., Sullivan, P.F., 2013. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* 45, 1150–1159. doi:10.1038/ng.2742
- Robling, S.A., Paykel, E.S., Dunn, V.J., Abbott, R., Katona, C., 2000. Long-term outcome of severe puerperal psychiatric illness: a 23 year follow-up study. *Psychol. Med.* 30, 1263–71.
- Roncero, C., Daigre, C., Gonzalvo, B., Valero, S., Castells, X., Grau-López, L., Eiroa-Orosa, F.J., Casas, M., 2013. Risk factors for cocaine-induced psychosis in cocaine-dependent patients. *Eur. Psychiatry* 28, 141–146. doi:https://doi.org/10.1016/j.eurpsy.2011.06.012
- Rooke, E.D.M., Vesterinen, H.M., Sena, E.S., Egan, K.J., Macleod, M.R., 2011. Dopamine agonists in animal models of Parkinson's disease: a systematic review and meta-analysis. *Parkinsonism Relat. Disord.* 17, 313–20. doi:10.1016/j.parkreldis.2011.02.010
- Rosenthal, R., 1979. The file drawer problem and tolerance for null results. *Psychol. Bull.* 86, 638–641. doi:10.1037/0033-2909.86.3.638
- Rosenthal, R., Fode, K.L., 1963. The effect of experimenter bias on the performance of the albino rat. *Behav. Sci.* 8, 183–189. doi:10.1002/bs.3830080302
- Roth, B., Hanizavareh, S.M., Blum, A., 2004. Serotonin receptors represent highly favorable molecular targets for cognitive enhancement in schizophrenia and other disorders. *Psychopharmacology (Berl.)* 174, 17–24. doi:10.1007/s00213-003-1683-8
- Ruth, J.A., Ullman, E.A., Collins, A.C., 1988. An analysis of cocaine effects on locomotor activities and heart rate in four inbred mouse strains. *Pharmacol. Biochem. Behav.* 29, 157–162. doi:10.1016/0091-3057(88)90289-4
- Sacco, K.A., Termine, A., Seyal, A., Dudas, M.M., Vessicchio, J.C., Krishnan-Sarin, S., Jatlow, P.I., Wexler, B.E., George, T.P., 2005. Effects of Cigarette Smoking on Spatial Working Memory and Attentional Deficits in Schizophrenia. *Arch. Gen. Psychiatry* 62, 649. doi:10.1001/archpsyc.62.6.649
- Sachdev, P., Keshavan, M.S., 2010. Secondary schizophrenia.
- Sahin, C., Doostdar, N., Neill, J.C., 2016. Towards the development of improved tests for negative symptoms of schizophrenia in a validated animal model. *Behav. Brain Res.* 312, 93–101. doi:10.1016/j.bbr.2016.06.021
- Samaranayake, S., 2006. Study quality in experimental stroke [WWW Document]. *Camarad. Monogr. No 2*. URL www.camarades.info/index_files/CM2.pdf
- Sampson, M., Barrowman, N.J., Moher, D., Clifford, T.J., Platt, R.W., Morrison, A.,

- Klassen, T.P., Zhang, L., 2006. Can electronic search engines optimize screening of search results in systematic reviews: an empirical study. *BMC Med. Res. Methodol.* 6, 7. doi:10.1186/1471-2288-6-7
- Samuel, G.O., Hoffmann, S., Wright, R.A., Lalu, M.M., Patlewicz, G., Becker, R.A., DeGeorge, G.L., Fergusson, D., Hartung, T., Lewis, R.J., Stephens, M.L., 2016. Guidance on assessing the methodological and reporting quality of toxicologically relevant studies: A scoping review. *Environ. Int.* 92–93, 630–646. doi:10.1016/j.envint.2016.03.010
- Sánchez-Morla, E.M., Mateo, J., Aparicio, A., García-Jiménez, M.Á., Jiménez, E., Santos, J.L., 2016. Prepulse inhibition in euthymic bipolar disorder patients in comparison with control subjects. *Acta Psychiatr. Scand.* 134, 350–359. doi:10.1111/acps.12604
- Sarnyai, Z., Jashar, C., Olivier, B., 2015. Modeling combined schizophrenia-related behavioral and metabolic phenotypes in rodents. *Behav. Brain Res.* 276, 130–42. doi:10.1016/j.bbr.2014.04.016
- Sarter, M., Nelson, C.L., Bruno, J.P., 2005. Cortical Cholinergic Transmission and Cortical Information Processing in Schizophrenia. *Schizophr. Bull.* 31, 117–138. doi:10.1093/schbul/sbi006
- Sato, M., Chen, C.C., Akiyama, K., Otsuki, S., 1983. Acute exacerbation of paranoid psychotic state after long-term abstinence in patients with previous methamphetamine psychosis. *Biol. Psychiatry* 18, 429–40.
- Savilla, K., Kettler, L., Galletly, C., 2008. Relationships Between Cognitive Deficits, Symptoms and Quality of Life in Schizophrenia. *Aust. New Zeal. J. Psychiatry* 42, 496–504. doi:10.1080/00048670802050512
- Schiffman, J., Abrahamson, A., Cannon, T., Labrie, J., Parnas, J., Schulsinger, F., Mednick, S., 2001. Early Rearing Factors in Schizophrenia. *Int. J. Ment. Health* 30, 3–16. doi:10.1080/00207411.2001.11449507
- Schizophrenia Working Group of the Psychiatric Genomics Consortium, S.W.G. of the P.G., Ripke, S., Neale, B.M., Corvin, A., Walters, J.T., Farh, K.-H., Holmans, P.A., Lee, P., Bulik-Sullivan, B., Collier, D.A., Huang, H., Pers, T.H., Agartz, I., Agerbo, E., Albus, M., Alexander, M., Amin, F., Bacanu, S.A., Begemann, M., Belliveau, R.A., Jr, Bene, J., Bergen, S.E., Bevilacqua, E., Bigdeli, T.B., Black, D.W., Bruggeman, R., Buccola, N.G., Buckner, R.L., Byerley, W., Cahn, W., Cai, G., Campion, D., Cantor, R.M., Carr, V.J., Carrera, N., Catts, S. V., Chambert, K.D., Chan, R.C., Chan, R.Y., Chen, E.Y., Cheng, W., Cheung, E.F., Chong, S.A., Cloninger, C.R., Cohen, D., Cohen, N., Cormican, P., Craddock, N., Crowley, J.J., Curtis, D., Davidson, M., Davis, K.L., Degenhardt, F., Favero, J. Del, Demontis, D., Dikeos, D., Dinan, T., Djurovic, S., Donohoe, G., Drapeau, E., Duan, J., Dudbridge, F., Durmishi, N., Eichhammer, P., Eriksson, J., Escott-Price, V., Essioux, L., Fanous, A.H., Farrell, M.S., Frank, J., Franke, L., Freedman, R., Freimer, N.B., Friedl, M., Friedman, J.I., Fromer, M., Genovese, G., Georgieva, L., Giegling, I., Giusti-Rodríguez, P., Godard, S., Goldstein, J.I., Golimbet, V., Gopal, S., Gratten, J., Haan, L. de, Hammer, C., Hamshere, M.L., Hansen, M., Hansen, T., Haroutunian, V., Hartmann, A.M., Henskens, F.A., Herms, S., Hirschhorn, J.N., Hoffmann, P., Hofman, A., Hollegaard, M. V, Hougaard, D.M., Ikeda, M., Joa,

I., Julià, A., Kahn, R.S., Kalaydjieva, L., Karachanak-Yankova, S., Karjalainen, J., Kavanagh, D., Keller, M.C., Kennedy, J.L., Khrunin, A., Kim, Y., Klovins, J., Knowles, J.A., Konte, B., Kucinskis, V., Kucinskiene, Z.A., Kuzelova-Ptackova, H., Kähler, A.K., Laurent, C., Lee, J., Lee, S.H., Legge, S.E., Lerer, B., Li, M., Li, T., Liang, K.-Y., Lieberman, J., Limborska, S., Loughland, C.M., Lubinski, J., Lönnqvist, J., Macek, M., Magnusson, P.K., Maher, B.S., Maier, W., Mallet, J., Marsal, S., Mattheisen, M., Mattingsdal, M., McCarley, R.W., McDonald, C., McIntosh, A.M., Meier, S., Meijer, C.J., Melegh, B., Melle, I., Meshulam-Gately, R.I., Metspalu, A., Michie, P.T., Milani, L., Milanova, V., Mokrab, Y., Morris, D.W., Mors, O., Murphy, K.C., Murray, R.M., Myin-Germeys, I., Müller-Myhsok, B., Nelis, M., Nenadic, I., Nertney, D.A., Nestadt, G., Nicodemus, K.K., Nikitina-Zake, L., Nisenbaum, L., Nordin, A., O'Callaghan, E., O'Dushlaine, C., O'Neill, F.A., Oh, S.-Y., Olincy, A., Olsen, L., Os, J. Van, Consortium, P.E.I., Pantelis, C., Papadimitriou, G.N., Papiol, S., Parkhomenko, E., Pato, M.T., Paunio, T., Pejovic-Milovancevic, M., Perkins, D.O., Pietiläinen, O., Pimm, J., Pocklington, A.J., Powell, J., Price, A., Pulver, A.E., Purcell, S.M., Quedest, D., Rasmussen, H.B., Reichenberg, A., Reimers, M.A., Richards, A.L., Roffman, J.L., Roussos, P., Ruderfer, D.M., Salomaa, V., Sanders, A.R., Schall, U., Schubert, C.R., Schulze, T.G., Schwab, S.G., Scolnick, E.M., Scott, R.J., Seidman, L.J., Shi, J., Sigurdsson, E., Silagadze, T., Silverman, J.M., Sim, K., Slominsky, P., Smoller, J.W., So, H.-C., Spencer, C.C.A., Stahl, E.A., Stefansson, H., Steinberg, S., Stogmann, E., Straub, R.E., Strengman, E., Strohmaier, J., Stroup, T.S., Subramaniam, M., Suvisaari, J., Svrakic, D.M., Szatkiewicz, J.P., Söderman, E., Thirumalai, S., Toncheva, D., Tosato, S., Veijola, J., Waddington, J., Walsh, D., Wang, D., Wang, Q., Webb, B.T., Weiser, M., Wildenauer, D.B., Williams, N.M., Williams, S., Witt, S.H., Wolen, A.R., Wong, E.H., Wormley, B.K., Xi, H.S., Zai, C.C., Zheng, X., Zimprich, F., Wray, N.R., Stefansson, K., Visscher, P.M., 2, W.T.C.-C.C., Adolfsson, R., Andreassen, O.A., Blackwood, D.H., Bramon, E., Buxbaum, J.D., Børglum, A.D., Cichon, S., Darvasi, A., Domenici, E., Ehrenreich, H., Esko, T., Gejman, P. V., Gill, M., Gurling, H., Hultman, C.M., Iwata, N., Jablensky, A. V., Jönsson, E.G., Kendler, K.S., Kirov, G., Knight, J., Lencz, T., Levinson, D.F., Li, Q.S., Liu, J., Malhotra, A.K., McCarroll, S.A., McQuillin, A., Moran, J.L., Mortensen, P.B., Mowry, B.J., Nöthen, M.M., Ophoff, R.A., Owen, M.J., Palotie, A., Pato, C.N., Petryshen, T.L., Posthuma, D., Rietschel, M., Riley, B.P., Rujescu, D., Sham, P.C., Sklar, P., Clair, D.S., Weinberger, D.R., Wendland, J.R., Werge, T., Daly, M.J., Sullivan, P.F., O'Donovan, M.C., 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–7. doi:10.1038/nature13595

Schulz, K.F., Altman, D.G., Moher, D., 2010. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 340, c332–c332. doi:10.1136/bmj.c332

Schulz, K.F., Chalmers, I., Hayes, R.J., Altman, D.G., 1995. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 273, 408–12.

Seeman, P., Lee, T., Chau-Wong, M., Wong, K., 1976. Antipsychotic drug doses and neuroleptic/dopamine receptors. *Nature* 261, 717–9.

Segal, D.S., Geyer, M.A., Schuckit, M.A., 1981. Stimulant-induced psychosis: an evaluation of animal methods. *Essays Neurochem. Neuropharmacol.* 5, 95–129.

- Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., Genovese, G., Rose, S.A., Handsaker, R.E., Daly, M.J., Carroll, M.C., Stevens, B., McCarroll, S.A., McCarroll, S.A., 2016. Schizophrenia risk from complex variation of complement component 4. *Nature* 530, 177–183. doi:10.1038/nature16549
- Selemon, L.D., Zecevic, N., 2015. Schizophrenia: a tale of two critical periods for prefrontal cortical development. *Transl. Psychiatry* 5, e623. doi:10.1038/tp.2015.115
- Semple, B.D., Blomgren, K., Gimlin, K., Ferriero, D.M., Noble-Haeusslein, L.J., 2013. Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. *Prog. Neurobiol.* 106–107, 1–16. doi:10.1016/j.pneurobio.2013.04.001
- Sena, E., van der Worp, H.B., Howells, D., Macleod, M., 2007. How can we improve the pre-clinical development of drugs for stroke? *Trends Neurosci.* 30, 433–9. doi:10.1016/j.tins.2007.06.009
- Sena, E.S., Currie, G.L., McCann, S.K., Macleod, M.R., Howells, D.W., 2014. Systematic reviews and meta-analysis of preclinical studies: why perform them and how to appraise them critically. *J. Cereb. Blood Flow Metab.* 34, 737–42. doi:10.1038/jcbfm.2014.28
- Sena, E.S., van der Worp, H.B., Bath, P.M.W., Howells, D.W., Macleod, M.R., 2010. Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy. *PLoS Biol.* 8, e1000344. doi:10.1371/journal.pbio.1000344
- Sengupta, P., 2013. The Laboratory Rat: Relating Its Age With Human's. *Int. J. Prev. Med.* 4, 624–30.
- Sesack, S.R., Carr, D.B., 2002. Selective prefrontal cortex inputs to dopamine cells: implications for schizophrenia. *Physiol. Behav.* 77, 513–517. doi:10.1016/S0031-9384(02)00931-9
- Shenton, M.E., Dickey, C.C., Frumin, M., McCarley, R.W., 2001. A review of MRI findings in schizophrenia. *Schizophr. Res.* 49, 1–52.
- Shipman, S.L., Astur, R.S., 2008. Factors affecting the hippocampal BOLD response during spatial memory. *Behav. Brain Res.* 187, 433–441. doi:10.1016/j.bbr.2007.10.014
- Shojania, K.G., Sampson, M., Ansari, M.T., Ji, J., Doucette, S., Moher, D., 2007. How Quickly Do Systematic Reviews Go Out of Date? A Survival Analysis. *Ann. Intern. Med.* 147, 224. doi:10.7326/0003-4819-147-4-200708210-00179
- Shuto, T., Nishi, A., 2011. Treatment of the Psychostimulant-Sensitized Animal Model of Schizophrenia. *CNS Neurosci. Ther.* 17, 133–139. doi:10.1111/j.1755-5949.2010.00218.x
- Silbersweig, D.A., Stern, E., Frith, C., Cahill, C., Holmes, A., Grootenok, S., Seaward, J., McKenna, P., Chua, S.E., Schnorr, L., Jones, T., Frackowiak, R.S.J., 1995. A functional neuroanatomy of hallucinations in schizophrenia. *Nature* 378, 176–179. doi:10.1038/378176a0

- Simpson, E.H., Winiger, V., Biezonski, D.K., Haq, I., Kandel, E.R., Kellendonk, C., 2014. Selective overexpression of dopamine D3 receptors in the striatum disrupts motivation but not cognition. *Biol. Psychiatry* 76, 823–31. doi:10.1016/j.biopsych.2013.11.023
- Sittig, L.J., Carbonetto, P., Engel, K.A., Krauss, K.S., Barrios-Camacho, C.M., Palmer, A.A., 2016. Genetic Background Limits Generalizability of Genotype-Phenotype Relationships. *Neuron* 91, 1253–1259. doi:10.1016/j.neuron.2016.08.013
- Smith, A.J., Clutton, R.E., Lilley, E., Hansen, K.E.A., Brattelid, T., 2017. PREPARE: guidelines for planning animal research and testing. *Lab. Anim.* 002367721772482. doi:10.1177/0023677217724823
- Smith, S., 2010. Gender differences in antipsychotic prescribing. *Int. Rev. Psychiatry* 22, 472–484. doi:10.3109/09540261.2010.515965
- Snitz, B.E., Macdonald, A.W., Carter, C.S., Carter, C.S., 2006. Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: a meta-analytic review of putative endophenotypes. *Schizophr. Bull.* 32, 179–94. doi:10.1093/schbul/sbi048
- Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* 45, 427–437. doi:10.1016/J.IPM.2009.03.002
- Stafford, M.R., Jackson, H., Mayo-Wilson, E., Morrison, A.P., Kendall, T., 2013. Early interventions to prevent psychosis: systematic review and meta-analysis. *BMJ* 346, f185. doi:10.1136/BMJ.F185
- Stahl, S.M., Muntner, N., 2013. *Stahl's essential psychopharmacology: neuroscientific basis and practical applications*. Cambridge University Press, Cambridge.
- Steeds, H., Carhart-Harris, R.L., Stone, J.M., 2015. Drug models of schizophrenia. *Ther. Adv. Psychopharmacol.* 5, 43–58. doi:10.1177/2045125314557797
- Stefanis, N., Frangou, S., Yakeley, J., Sharma, T., O'Connell, P., Morgan, K., Sigmundsson, T., Taylor, M., Murray, R., 1999. Hippocampal volume reduction in schizophrenia: effects of genetic risk and pregnancy and birth complications. *Biol. Psychiatry* 46, 697–702.
- Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O.P.H., Mors, O., Mortensen, P.B., Sigurdsson, E., Gustafsson, O., Nyegaard, M., Tuulio-Henriksson, A., Ingason, A., Hansen, T., Suvisaari, J., Lonnqvist, J., Paunio, T., Børglum, A.D., Hartmann, A., Fink-Jensen, A., Nordentoft, M., Hougaard, D., Norgaard-Pedersen, B., Böttcher, Y., Olesen, J., Breuer, R., Möller, H.-J., Giegling, I., Rasmussen, H.B., Timm, S., Mattheisen, M., Bitter, I., Réthelyi, J.M., Magnusdottir, B.B., Sigmundsson, T., Olason, P., Masson, G., Gulcher, J.R., Haraldsson, M., Fossdal, R., Thorgeirsson, T.E., Thorsteinsdottir, U., Ruggieri, M., Tosato, S., Franke, B., Strengman, E., Kiemenev, L.A., GROUP†, Melle, I., Djurovic, S., Abramova, L., Kaleda, V., Sanjuan, J., de Frutos, R., Bramon, E., Vassos, E., Fraser, G., Etinger, U., Picchioni, M., Walker, N., Toulopoulou, T., Need, A.C., Ge, D., Lim Yoon, J., Shianna, K. V., Freimer, N.B., Cantor, R.M., Murray, R., Kong, A.,

- Golimbet, V., Carracedo, A., Arango, C., Costas, J., Jönsson, E.G., Terenius, L., Agartz, I., Petursson, H., Nöthen, M.M., Rietschel, M., Matthews, P.M., Muglia, P., Peltonen, L., St Clair, D., Goldstein, D.B., Stefansson, K., Collier, D.A., Kahn, R.S., Linszen, D.H., van Os, J., Wiersma, D., Bruggeman, R., Cahn, W., de Haan, L., Krabbendam, L., Myin-Germeys, I., 2009. Common variants conferring risk of schizophrenia. *Nature* 460, 744–7. doi:10.1038/nature08186
- Steinmetz, K.L., Spack, E.G., 2009. The basics of preclinical drug development for neurodegenerative disease indications. *BMC Neurol.* 9, S2. doi:10.1186/1471-2377-9-S1-S2
- Stern, T.A., Rosenbaum, J.F., Fava, M., Biederman, J., Rauch, S.L., 2008. *Massachusetts General Hospital Comprehensive Clinical Psychiatry*. Elsevier Health Sciences.
- Stevinson, C., Lawlor, D.A., 2004. Searching multiple databases for systematic reviews: added value or diminishing returns? *Complement. Ther. Med.* 12, 228–232. doi:10.1016/j.ctim.2004.09.003
- Steward, O., 2016. A Rhumba of “R’s” Replication, Reproducibility, Rigor, Robustness: What Does a Failure to Replicate Mean? *eNeuro* 3. doi:10.1523/ENEURO.0072-16.2016
- Stokes, P.R.A., Shotbolt, P., Mehta, M.A., Turkheimer, E., Benecke, A., Copeland, C., Turkheimer, F.E., Lingford-Hughes, A.R., Howes, O.D., 2013. Nature or Nurture? Determining the Heritability of Human Striatal Dopamine Function: an [18F]-DOPA PET Study. *Neuropsychopharmacology* 38, 485–491. doi:10.1038/npp.2012.207
- Sullivan, P.F., Kendler, K.S., Neale, M.C., 2003. Schizophrenia as a Complex Trait. *Arch. Gen. Psychiatry* 60, 1187. doi:10.1001/archpsyc.60.12.1187
- Swerdlow, N.R., Light, G.A., Cadenhead, K.S., Sprock, J., Hsieh, M.H., Braff, D.L., 2006. Startle Gating Deficits in a Large Cohort of Patients With Schizophrenia. *Arch. Gen. Psychiatry* 63, 1325–35. doi:10.1001/archpsyc.63.12.1325
- Swinney, D.C., Anthony, J., 2011. How were new medicines discovered? *Nat. Rev. Drug Discov.* 10, 507–519. doi:10.1038/nrd3480
- Takahashi, H., Hashimoto, R., Iwase, M., Ishii, R., Kamio, Y., Takeda, M., 2011. Prepulse inhibition of startle response: recent advances in human studies of psychiatric disease. *Clin. Psychopharmacol. Neurosci.* 9, 102–10. doi:10.9758/cpn.2011.9.3.102
- Tandon, R., 1999. Moving beyond findings: Concepts and model-building in schizophrenia. *J. Psychiatr. Res.* doi:10.1016/S0022-3956(99)00036-9
- Tandon, R., Keshavan, M.S., Nasrallah, H.A., 2008. Schizophrenia, “Just the Facts”: What we know in 2008. *Schizophr. Res.* 100, 4–19. doi:10.1016/j.schres.2008.01.022
- Tatem, K.S., Quinn, J.L., Phadke, A., Yu, Q., Gordish-Dressman, H., Nagaraju, K., 2014. Behavioral and locomotor measurements using an open field activity monitoring system for skeletal muscle diseases. *J. Vis. Exp.* 51785.

doi:10.3791/51785

ter Riet, G., Korevaar, D.A., Leenaars, M., Sterk, P.J., Van Noorden, C.J.F., Bouter, L.M., Lutter, R., Elferink, R.P.O., Hooft, L., 2012. Publication Bias in Laboratory Animal Research: A Survey on Magnitude, Drivers, Consequences and Potential Solutions. *PLoS One* 7, e43404. doi:10.1371/journal.pone.0043404

Terry, A. V, Jr, 2008. Role of the central cholinergic system in the therapeutics of schizophrenia. *Curr. Neuropharmacol.* 6, 286–92. doi:10.2174/157015908785777247

Thomsen, M., Caine, S.B., 2011. Psychomotor stimulant effects of cocaine in rats and 15 mouse strains. *Exp. Clin. Psychopharmacol.* 19, 321–41. doi:10.1037/a0024798

Tienari, P., 1991. Interaction between genetic vulnerability and family environment: the Finnish adoptive family study of schizophrenia. *Acta Psychiatr. Scand.* 84, 460–465. doi:10.1111/j.1600-0447.1991.tb03178.x

Tsilidis, K.K., Panagiotou, O.A., Sena, E.S., Aretouli, E., Evangelou, E., Howells, D.W., Salaman, R.A.-S., Macleod, M.R., Ioannidis, J.P.A., 2013. Evaluation of Excess Significance Bias in Animal Studies of Neurological Diseases. *PLoS Biol.* 11, e1001609. doi:10.1371/journal.pbio.1001609

Ujike, H., 2002. Stimulant-induced psychosis and schizophrenia: The role of sensitization. *Curr. Psychiatry Rep.* 4, 177–184. doi:10.1007/s11920-002-0024-7

United Nations Office on Drugs and Crime, 2016. United Nations Office on Drugs and Crime : UNODC.

Valsamis, B., Schmid, S., 2011. Habituation and prepulse inhibition of acoustic startle in rodents. *J. Vis. Exp.* e3446. doi:10.3791/3446

van den Buuse, M., 2010. Modeling the positive symptoms of schizophrenia in genetically modified mice: pharmacology and methodology aspects. *Schizophr. Bull.* 36, 246–70. doi:10.1093/schbul/sbp132

van der Staay, F.J., Arndt, S.S., Nordquist, R.E., 2009. Evaluation of animal models of neurobehavioral disorders. *Behav. Brain Funct.* 5, 11. doi:10.1186/1744-9081-5-11

van der Worp, H.B., Howells, D.W., Sena, E.S., Porritt, M.J., Rewell, S., O'Collins, V., Macleod, M.R., Hackam, D., Redelmeier, D., Perel, P., Roberts, I., Sena, E., Wheble, P., Briscoe, C., Hackam, D., Besselink, M., Santvoort, H. van, Buskens, E., Boermeester, M., Goor, H. van, Shuaib, A., Lees, K., Lyden, P., Grotta, J., Davalos, A., Dirksen, M., Laarman, G., Simoons, M., Duncker, D., Yellon, D., Hausenloy, D., O'Collins, V., Macleod, M., Donnan, G., Horky, L., Worp, B. van der, Worp, H. Van der, Gijn, J. Van, Grotta, J., Gladstone, D., Black, S., Hakim, A., Juni, P., Altman, D., Egger, M., Altman, D., Schulz, K., Moher, D., Egger, M., Davidoff, F., Dirnagl, U., Sena, E., Worp, H. Van der, Howells, D., Macleod, M., Macleod, M., Fisher, M., O'Collins, V., Sena, E., Dirnagl, U., Rosenthal, R., Worp, H. Van der, Haan, P. de, Morrema, E., Kalkman, C., Worp, H. Van der, Sena, E., Donnan, G., Howells, D., Macleod, M., Campbell, M., Julious, S., Altman, D., Festing, M., Altman, D., Bath, P.,

- Macleod, M., Green, A., Lee, D., Nguyen, Q., Lapointe, N., Austin, P., Ohlsson, A., Macleod, M., Worp, H. Van der, Sena, E., Howells, D., Dirnagl, U., Samaranyake, S., Miettinen, O., Schulz, K., Chalmers, I., Hayes, R., Altman, D., Noseworthy, J., ebers, G., Vandervoort, M., Farquhar, R., Yetisir, E., Schulz, K., Grimes, D., Schulz, K., Grimes, D., Schulz, K., Grimes, D., Bebart, V., Luyten, D., Heard, K., Macleod, M., O'Collins, T., Howells, D., Donnan, G., Crossley, N., Sena, E., Goehler, J., Horn, J., der, W. van, Worp, H. Van der, Raaijmakers, T., Kappelle, L., Saver, J., Johnston, K., Homer, D., Wityk, R., Koroshetz, W., Kimmelman, J., London, A., Ravina, B., Ramsay, T., Bernstein, M., Scott, S., Kranz, J., Cole, J., Lincecum, J., Thompson, K., Dwan, K., Altman, D., Arnaiz, J., Bloom, J., Chan, A., Dirx, M., Zeegers, M., Dagnelie, P., Bogaard, T. van den, Brandt, P. van den, Macleod, M., O'Collins, T., Horky, L., Howells, D., Donnan, G., Juutilainen, J., Kumlin, T., Naarala, J., Benatar, M., Neitzke, U., Harder, T., Schellong, K., Melchior, K., Ziska, T., Sena, E., Worp, H. Van der, Bath, P., Howells, D., Macleod, M., Angelis, C. De, Drazen, J., Frizelle, F., Haug, C., Hoey, J., Willmot, M., Gibson, C., Gray, L., Murphy, S., Bath, P., Willmot, M., Gray, L., Gibson, C., Murphy, S., Bath, P., Gibson, C., Gray, L., Murphy, S., Bath, P., Gibson, C., Gray, L., Bath, P., Murphy, S., Banwell, V., Sena, E., Macleod, M., 2010. Can Animal Models of Disease Reliably Inform Human Studies? *PLoS Med.* 7, e1000245. doi:10.1371/journal.pmed.1000245
- van der Worp, H.B., Sena, E.S., Donnan, G.A., Howells, D.W., Macleod, M.R., 2007. Hypothermia in animal models of acute ischaemic stroke: a systematic review and meta-analysis. *Brain* 130, 3063–74. doi:10.1093/brain/awm083
- Van Norman, G.A., 2016. Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs, *JACC: Basic to Translational Science*. doi:10.1016/j.jacbts.2016.03.002
- van Oel, C.J., Sitskoorn, M.M., Cremer, M.P.M., Kahn, R.S., 2002. School Performance as a Premorbid Marker for Schizophrenia: A Twin Study. *Schizophr. Bull.* 28, 401–414. doi:10.1093/oxfordjournals.schbul.a006949
- van Os, J., Kapur, S., 2009. Schizophrenia. *Lancet* 374, 635–45. doi:10.1016/S0140-6736(09)60995-8
- Varga, O.E., Hansen, A.K., Sandøe, P., Anna, I., Olsson, S., 2010. Comment: Validating animal models for preclinical research: a scientific and ethical discussion. *ATLA* 38, 245–248.
- Vassos, E., Pedersen, C.B., Murray, R.M., Collier, D.A., Lewis, C.M., 2012. Meta-analysis of the association of urbanicity with schizophrenia. *Schizophr. Bull.* 38, 1118–23. doi:10.1093/schbul/sbs096
- Vergara-Moragues, E., Araos Gómez, P., González-Saiz, F., Rodríguez-Fonseca, F., 2014. Cocaine-induced psychotic symptoms in clinical setting. *Psychiatry Res.* 217, 115–120. doi:10.1016/j.psychres.2014.02.024
- Vesterinen, H.M., Sena, E.S., Egan, K.J., Hirst, T.C., Churolov, L., Currie, G.L., Antonic, A., Howells, D.W., Macleod, M.R., 2014. Meta-analysis of data from animal studies: a practical guide. *J. Neurosci. Methods* 221, 92–102. doi:10.1016/j.jneumeth.2013.09.010

- Vesterinen, H.M., Sena, E.S., French-Constant, C., Williams, A., Chandran, S., Macleod, M.R., 2010. Improving the translational hit of experimental treatments in multiple sclerosis. *Mult. Scler.* 16, 1044–55. doi:10.1177/1352458510379612
- Vita, A., De Peri, L., Silenzi, C., Dieci, M., 2006. Brain morphology in first-episode schizophrenia: A meta-analysis of quantitative magnetic resonance imaging studies. *Schizophr. Res.* 82, 75–88. doi:10.1016/j.schres.2005.11.004
- Vreeker, A., van Bergen, A.H., Kahn, R.S., 2015. Cognitive enhancing agents in schizophrenia and bipolar disorder. *Eur. Neuropsychopharmacol.* 25, 969–1002. doi:10.1016/J.EURONEURO.2015.04.014
- Wahlsten, D., Metten, P., Phillips, T.J., Boehm, S.L., Burkhart-Kasch, S., Dorow, J., Doerksen, S., Downing, C., Fogarty, J., Rodd-Henricks, K., Hen, R., McKinnon, C.S., Merrill, C.M., Nolte, C., Schalomon, M., Schlumbohm, J.P., Sibert, J.R., Wenger, C.D., Dudek, B.C., Crabbe, J.C., 2003. Different data from different labs: Lessons from studies of gene-environment interaction. *J. Neurobiol.* 54, 283–311. doi:10.1002/neu.10173
- Wang, Q., Liao, J., Hair, K., Bannach-Brown, A., Bahor, Z., Currie, G.L., McCann, S.K., Howells, D.W., Sena, E.S., Macleod, M.R., 2018. Estimating the statistical performance of different approaches to meta-analysis of data from animal studies in identifying the impact of aspects of study design. *bioRxiv* 256776. doi:10.1101/256776
- Weickert, C.S., Weickert, T.W., Pillai, A., Buckley, P.F., 2013. Biomarkers in Schizophrenia: A Brief Conceptual Consideration. *Dis. Markers* 35, 3–9. doi:10.1155/2013/510402
- Weidenauer, A., Bauer, M., Sauerzopf, U., Bartova, L., Praschak-Rieder, N., Sitte, H.H., Kasper, S., Willeit, M., 2017. Making Sense of: Sensitization in Schizophrenia. *Int. J. Neuropsychopharmacol.* 20, 1–10. doi:10.1093/ijnp/pyw081
- Weinberger, D.R., 2013. Diagnostic criteria: Don't oversimplify psychiatric disorders. *Nature* 497, 565–565. doi:10.1038/497565d
- Weinberger, D.R., Gallhofer, B., 1997. Cognitive function in schizophrenia. *Int. Clin. Psychopharmacol.* 12 Suppl 4, S29-36.
- Wells, B.G., Marken, P.A., Rickman, L.A., Brown, C.S., Hamann, G., Grimmig, J., 1989. Characterizing anticholinergic abuse in community mental health. *J. Clin. Psychopharmacol.* 9, 431–5.
- Whiting, P., Savović, J., Higgins, J.P.T., Caldwell, D.M., Reeves, B.C., Shea, B., Davies, P., Kleijnen, J., Churchill, R., ROBIS group, R., 2016. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J. Clin. Epidemiol.* 69, 225–34. doi:10.1016/j.jclinepi.2015.06.005
- Wieschowski, S., Silva, D.S., Streh, D., 2016. Animal Study Registries: Results from a Stakeholder Analysis on Potential Strengths, Weaknesses, Facilitators, and Barriers. *PLOS Biol.* 14, e2000391. doi:10.1371/journal.pbio.2000391
- Willner, P., 1984. The validity of animal models of depression. *Psychopharmacology (Berl.)* 83, 1–16. doi:10.1007/BF00427414

- Wilson, C., Terry, A. V., Jr., 2010. Neurodevelopmental animal models of schizophrenia: role in novel drug discovery and development. *Clin. Schizophr. Relat. Psychoses* 4, 124–37. doi:10.3371/CSRP.4.2.4
- Wilson, C.A., Koenig, J.I., 2014. Social interaction and social withdrawal in rodents as readouts for investigating the negative symptoms of schizophrenia. *Eur. Neuropsychopharmacol.* 24, 759–73. doi:10.1016/j.euroneuro.2013.11.008
- Woods, B.T., Ward, K.E., Johnson, E.H., 2005. Meta-analysis of the time-course of brain volume reduction in schizophrenia: implications for pathogenesis and early treatment. *Schizophr. Res.* 73, 221–228. doi:10.1016/j.schres.2004.05.014
- World Health Organization, 2016. ICD-10 : international statistical classification of diseases and related health problems.
- World Health Organization, 2001. The world health report 2001 : mental health : new understanding, new hope.
- Wright, I.C., Rabe-Hesketh, S., Woodruff, P.W.R., David, A.S., Murray, R.M., Bullmore, E.T., 2000. Meta-Analysis of Regional Brain Volumes in Schizophrenia. *Am. J. Psychiatry* 157, 16–25. doi:10.1176/ajp.157.1.16
- Yang, Y., Marder, S., Green, M., 2017. Repurposing Drugs for Cognition in Schizophrenia. *Clin. Pharmacol. Ther.* 101, 191–193. doi:10.1002/cpt.529
- Young, J.W., Markou, A., 2015. Translational Rodent Paradigms to Investigate Neuromechanisms Underlying Behaviors Relevant to Amotivation and Altered Reward Processing in Schizophrenia. *Schizophr. Bull.* 41, 1024–34. doi:10.1093/schbul/sbv093
- Young, J.W., Powell, S.B., Risbrough, V., Marston, H.M., Geyer, M.A., 2009. Using the MATRICS to guide development of a preclinical cognitive test battery for research in schizophrenia. *Pharmacol. Ther.* 122, 150–202. doi:10.1016/j.pharmthera.2009.02.004
- Zajackowski, W., Czyrak, A., Wedzony, K., 2003. A competitive antagonist of NMDA receptors CGP 40116 attenuates experimental symptoms of schizophrenia evoked by MK-801. *Pol. J. Pharmacol.* 55, 703–11.
- Zhang, Z.J., Reynolds, G.P., 2002. A selective decrease in the relative density of parvalbumin-immunoreactive neurons in the hippocampus in schizophrenia. *Schizophr. Res.* 55, 1–10.
- Ziermans, T., Schothorst, P., Magnée, M., van Engeland, H., Kemner, C., 2011. Reduced prepulse inhibition in adolescents at risk for psychosis: a 2-year follow-up study. *J. Psychiatry Neurosci.* 36, 127–34. doi:10.1503/jpn.100063
- Zink, M., Correll, C.U., 2015. Glutamatergic agents for schizophrenia: current evidence and perspectives. *Expert Rev. Clin. Pharmacol.* 8, 335–352. doi:10.1586/17512433.2015.1040393
- Zombeck, J.A., Swearingen, S.P., Rhodes, J.S., 2010. Acute locomotor responses to cocaine in adolescents vs. adults from four divergent inbred mouse strains. *Genes, Brain Behav.* 9, 892–898. doi:10.1111/j.1601-183X.2010.00630.x

Zwetsloot, P.-P., Van Der Naald, M., Sena, E.S., Howells, D.W., IntHout, J., De Groot, J.A., Chamuleau, S.A., MacLeod, M.R., Wever, K.E., 2017. Standardized mean differences cause funnel plot distortion in publication bias assessments. *Elife* 6, e24260. doi:10.7554/eLife.24260

psychotherapiekontrolle[All Fields] OR psychotherapiekonzept[All Fields] OR
psychotherapiemethoden[All Fields] OR psychotherapiemotivation[All Fields] OR
psychotherapien[All Fields] OR psychotherapiepatienten[All Fields] OR
psychotherapiepatientinnen[All Fields] OR psychotherapieplanung[All Fields] OR
psychotherapieplatz[All Fields] OR psychotherapiepraxis[All Fields] OR
psychotherapieprinzip[All Fields] OR psychotherapieprogramme[All Fields] OR
psychotherapieprozessforschung[All Fields] OR psychotherapierelevanten[All Fields] OR
psychotherapierichtlinien[All Fields] OR psychotherapies[All Fields] OR psychotherapies'[All
Fields] OR psychotherapieschulen[All Fields] OR psychotherapiestation[All Fields] OR
psychotherapiestationen[All Fields] OR psychotherapiestudie[All Fields] OR
psychotherapiestudien[All Fields] OR psychotherapieverfahren[All Fields] OR
psychotherapieverlauf[All Fields] OR psychotherapieverlaufe[All Fields] OR
psychotherapieverlaufsforschung[All Fields] OR psychotherapieversuch[All Fields] OR
psychotherapiewoche[All Fields] OR psychotherapii[All Fields] OR psychotherapique[All
Fields] OR psychotherapique[All Fields] OR psychotherapiques[All Fields] OR
psychotherapist[All Fields] OR psychotherapist'[All Fields] OR psychotherapist's[All Fields]
OR psychotherapiste[All Fields] OR psychotherapists[All Fields] OR psychotherapists'[All
Fields] OR psychotherapits[All Fields] OR psychotherapty[All Fields] OR
psychotherapuetic[All Fields] OR psychotheraputic[All Fields] OR psychotherapy[All Fields]
OR psychotherapy[All Fields] OR psychotherapy's[All Fields] OR psychotherapy1[All Fields]
OR psychotherapygenetics[All Fields] OR psychotherapyuniversity[All Fields] OR
psychotherapyw2[All Fields] OR psychotheratpie[All Fields] OR psychotheray[All Fields] OR
psychotherayp[All Fields] OR psychothereutic[All Fields] OR psychotheroapy[All Fields] OR
psychotherories[All Fields] OR psychotherpay[All Fields] OR psychotherpeutic[All Fields] OR
psychotherpie[All Fields] OR psychotherpiemotivation[All Fields] OR psychotherpists[All
Fields] OR psychotherpy[All Fields] OR psychotherrapeutic[All Fields] OR psychotherspy[All
Fields] OR psychothrapeutic[All Fields] OR psychothraphy[All Fields] OR psychothrapie[All
Fields] OR psychothropic[All Fields] OR psychothymic[All Fields] OR psychotic[All Fields]
OR psychotic'[All Fields] OR psychotic's[All Fields] OR psychotica[All Fields] OR
psychotical[All Fields] OR psychotically[All Fields] OR psychoticbody[All Fields] OR
psychotici[All Fields] OR psychoticism[All Fields] OR psychoticism'[All Fields] OR
psychoticismu[All Fields] OR psychoticity[All Fields] OR psychoticity'[All Fields] OR
psychoticzim[All Fields] OR psychotick'y[All Fields] OR psychotick'ym[All Fields] OR
psychoticka[All Fields] OR psychoticke[All Fields] OR psychotickeho[All Fields] OR
psychoticky[All Fields] OR psychotickych[All Fields] OR psychoticlke[All Fields] OR
psychoticng[All Fields] OR psychotics[All Fields] OR psychotics'[All Fields] OR
psychoticum[All Fields] OR psychotiform[All Fields] OR psychotiform'[All Fields] OR
psychotigues[All Fields] OR psychotiker[All Fields] OR psychotikern[All Fields] OR
psychotikmi[All Fields] OR psychotiku[All Fields] OR psychotikum[All Fields] OR

psychotikus[All Fields] OR psychotikusokon[All Fields] OR psychotiky[All Fields] OR
psychotique[All Fields] OR psychotiques[All Fields] OR psychotisation[All Fields] OR
psychotisch[All Fields] OR psychotische[All Fields] OR psychotische'[All Fields] OR
psychotischem[All Fields] OR psychotischen[All Fields] OR psychotischer[All Fields] OR
psychotisches[All Fields] OR psychoticism[All Fields] OR psychotism[All Fields] OR
psychotizing[All Fields] OR psychotizismus[All Fields] OR psychoto[All Fields] OR
psychotocism[All Fields] OR psychotoform[All Fields] OR psychotogen[All Fields] OR
psychotogenesis[All Fields] OR psychotogenic[All Fields] OR psychotogenicity[All Fields] OR
psychotogens[All Fields] OR psychotogique[All Fields] OR psychotogiques[All Fields] OR
psychotoid[All Fields] OR psychotoid'[All Fields] OR psychotoksykologii[All Fields] OR
psycholysis[All Fields] OR psycholytic[All Fields] OR psychotomimesis[All Fields] OR
psychotomimetic[All Fields] OR psychotomimetic'[All Fields] OR psychotomimetically[All
Fields] OR psychotomimetics[All Fields] OR psychotomimeticumok[All Fields] OR
psychotomimetika[All Fields] OR psychotomimétique[All Fields] OR psychotomimétiques[All
Fields] OR psychotomimétique[All Fields] OR psychotomimetric[All Fields] OR
psychotomimetic[All Fields] OR psychotomorphic[All Fields] OR psychotomymetic[All
Fields] OR psychoton[All Fields] OR psychotone[All Fields] OR psychotonic[All Fields] OR
psychotonica[All Fields] OR psychotonics[All Fields] OR psychotonicumok[All Fields] OR
psychotonika[All Fields] OR psychotonikum[All Fields] OR psychotonikums[All Fields] OR
psychotonin[All Fields] OR psychotonique[All Fields] OR psychotoniques[All Fields] OR
psychotonisierender[All Fields] OR psychotonovy[All Fields] OR psychotonu[All Fields] OR
psychotop[All Fields] OR psychotopes[All Fields] OR psychotopic[All Fields] OR
psychotopics[All Fields] OR psychotorpnijch[All Fields] OR psychototmimetic[All Fields] OR
psychotoxic[All Fields] OR psychotoxical[All Fields] OR psychotoxicity[All Fields] OR
psychotoxicologic[All Fields] OR psychotoxicological[All Fields] OR psychotoxicology[All
Fields] OR psychotoxikologie[All Fields] OR psychotoxikologische[All Fields] OR
psychotoxin[All Fields] OR psychotoxins[All Fields] OR psychotoxischen[All Fields] OR
psychotraining[All Fields] OR psychotramide[All Fields] OR psychotrauma[All Fields] OR
psychotrauma's[All Fields] OR psychotraumagroep[All Fields] OR psychotraumas[All Fields]
OR psychotraumas'[All Fields] OR psychotraumata[All Fields] OR psychotraumatic[All Fields]
OR psychotraumatique[All Fields] OR psychotraumatiques[All Fields] OR
psychotraumatischen[All Fields] OR psychotraumatised[All Fields] OR psychotraumatism[All
Fields] OR psychotraumatisme[All Fields] OR psychotraumatismes[All Fields] OR
psychotraumatitiserter[All Fields] OR psychotraumatizace[All Fields] OR
psychotraumatization[All Fields] OR psychotraumatized[All Fields] OR
psychotraumatizing[All Fields] OR psychotraumatol[All Fields] OR psychotraumatologi[All
Fields] OR psychotraumatological[All Fields] OR psychotraumatologie[All Fields] OR
psychotraumatologische[All Fields] OR psychotraumatologischen[All Fields] OR
psychotraumatologischer[All Fields] OR psychotraumatology[All Fields] OR psychotria[All

Fields] OR psychotriae[All Fields] OR psychotrianoside[All Fields] OR psychotrianosides[All Fields] OR psychotrichological[All Fields] OR psychotrichology[All Fields] OR psychotridine[All Fields] OR psychotrieae[All Fields] OR psychotriifolia[All Fields] OR psychotrimine[All Fields] OR psychotrine[All Fields] OR psychotripine[All Fields] OR psychotroic[All Fields] OR psychotronic[All Fields] OR psychotronice[All Fields] OR psychotronics[All Fields] OR psychotronika[All Fields] OR psychotrop[All Fields] OR psychotropa[All Fields] OR psychotrope[All Fields] OR psychotropen[All Fields] OR psychotroper[All Fields] OR psychotropes[All Fields] OR psychotroph[All Fields] OR psychotropes[All Fields] OR psychotrophic[All Fields] OR psychotropics[All Fields] OR psychotrophie[All Fields] OR psychotrophs[All Fields] OR psychotrophia[All Fields] OR psychotropic[All Fields] OR psychotropic'[All Fields] OR psychotropic's[All Fields] OR psychotropica[All Fields] OR psychotropical[All Fields] OR psychotropically[All Fields] OR psychotropicity[All Fields] OR psychotropicmedication[All Fields] OR psychotropics[All Fields] OR psychotropics'[All Fields] OR psychotropid[All Fields] OR psychotropism[All Fields] OR psychotropn'ich[All Fields] OR psychotropne[All Fields] OR psychotropneho[All Fields] OR psychotropni[All Fields] OR psychotropnich[All Fields] OR psychotropnimi[All Fields] OR psychotropowe[All Fields] OR psychotropowego[All Fields] OR psychotropowych[All Fields] OR psychotropowym[All Fields] OR psychotropowymi[All Fields] OR psychotrops[All Fields] OR psychotropy[All Fields] OR psychotroropes[All Fields] OR psychotrpic[All Fields] OR psychotuberculosis[All Fields] OR psychotychnych[All Fields] OR psychotyczna[All Fields] OR psychotyczne[All Fields] OR psychotycznego[All Fields] OR psychotycznej[All Fields] OR psychotyczny[All Fields] OR psychotycznych[All Fields] OR psychotycznym[All Fields] OR psychotycznymi[All Fields] OR psychotycznych[All Fields] OR psychotype[All Fields] OR psychotypes[All Fields] OR psychotypical[All Fields] OR psychotypological[All Fields] OR psychotypology[All Fields] OR psychotycznymi[All Fields] OR psychotyzm[All Fields]) OR ("psychotic disorders"[MeSH Terms] OR ("psychotic"[All Fields] AND "disorders"[All Fields]) OR "psychotic disorders"[All Fields] OR "psychosis"[All Fields])) OR ("psychotic disorders"[MeSH Terms] OR ("psychotic"[All Fields] AND "disorders"[All Fields]) OR "psychotic disorders"[All Fields] OR "psychoses"[All Fields])) OR ("paranoid disorders"[MeSH Terms] OR ("paranoid"[All Fields] AND "disorders"[All Fields]) OR "paranoid disorders"[All Fields] OR "paranoia"[All Fields])) OR paraphrenia[All Fields] OR (sensitive[All Fields] AND beziehungswahn[All Fields])) OR (involuntional[All Fields] AND paranoid[All Fields] AND state[All Fields])) OR ("shared paranoid disorder"[MeSH Terms] OR ("shared"[All Fields] AND "paranoid"[All Fields] AND "disorder"[All Fields]) OR "shared paranoid disorder"[All Fields] OR ("folie"[All Fields] AND "deux"[All Fields]) OR "folie deux"[All Fields])) OR (catatonia[All Fields] OR catatonia'[All Fields] OR catatonialike[All Fields] OR catatonias[All Fields] OR catatonic[All Fields] OR catatonica[All Fields] OR catatonical[All Fields] OR catatoniche[All Fields] OR catatonici[All Fields] OR catatoniclike[All Fields] OR catatonico[All Fields] OR catatonicocatalettiche[All Fields] OR catatonicos[All

Fields] OR catatonics[All Fields] OR catatonie[All Fields] OR catatonies[All Fields] OR
catatoniform[All Fields] OR catatoniformpsychomotor[All Fields] OR catatonig'ene[All Fields]
OR catatonigene[All Fields] OR catatonigenes[All Fields] OR catatonigenic[All Fields] OR
catatonin[All Fields] OR catatonine[All Fields] OR catatonique[All Fields] OR catatoniques[All
Fields] OR catatonis[All Fields] OR catatonizante[All Fields] OR catatonizzati[All Fields] OR
catatono[All Fields] OR catatonogenic[All Fields] OR catatonoid[All Fields] OR catatony[All
Fields])) OR (delusion[All Fields] OR delusion'[All Fields] OR delusion's[All Fields] OR
delusional[All Fields] OR delusional'[All Fields] OR delusionalilty[All Fields] OR
delusionalism[All Fields] OR delusionality[All Fields] OR delusionally[All Fields] OR
delusionals[All Fields] OR delusionary[All Fields] OR delusione[All Fields] OR delusiones[All
Fields] OR delusioni[All Fields] OR delusionlike[All Fields] OR delusionnal[All Fields] OR
delusions[All Fields] OR delusions'[All Fields])) OR (hallucinate[All Fields] OR
hallucinated[All Fields] OR hallucinates[All Fields] OR hallucinatic[All Fields] OR
hallucinatie[All Fields] OR hallucinaties[All Fields] OR hallucinatiion[All Fields] OR
hallucinatiions[All Fields] OR hallucinating[All Fields] OR hallucination[All Fields] OR
hallucination'[All Fields] OR hallucinational[All Fields] OR hallucinationen[All Fields] OR
hallucinationer[All Fields] OR hallucinationes[All Fields] OR hallucinations[All Fields] OR
hallucinations'[All Fields] OR hallucinations'than[All Fields] OR hallucinatire[All Fields] OR
hallucinative[All Fields] OR hallucinatoire[All Fields] OR hallucinatoire'[All Fields] OR
hallucinatoires[All Fields] OR hallucinatoiresaigues[All Fields] OR hallucinator[All Fields] OR
hallucinatoria[All Fields] OR hallucinatoric[All Fields] OR hallucinatornih[All Fields] OR
hallucinators[All Fields] OR hallucinators'[All Fields] OR hallucinatory[All Fields] OR
hallucinatory'[All Fields] OR hallucinatroy[All Fields])) OR (schizotypai[All Fields] OR
schizotypal[All Fields] OR schizotypal'[All Fields] OR schizotypal"[All Fields] OR
schizotypality[All Fields] OR schizotypals[All Fields] OR schizotype[All Fields] OR
schizotype's[All Fields] OR schizotypes[All Fields] OR schizotypes'[All Fields] OR
schizotypia[All Fields] OR schizotypic[All Fields] OR schizotypical[All Fields] OR
schizotypics[All Fields] OR schizotypie[All Fields] OR schizotypies[All Fields] OR
schizotypique[All Fields] OR schizotypische[All Fields] OR schizotypischen[All Fields] OR
schizotypital[All Fields] OR schizotypy[All Fields] OR schizotypy'[All Fields] OR
schizotypya[All Fields] OR schizotypys[All Fields])) OR psychoactive[All Fields]) OR
oneirophrenia[All Fields]) OR (psychogen[All Fields] OR psychogender[All Fields] OR
psychogene[All Fields] OR psychogeneic[All Fields] OR psychogenem[All Fields] OR
psychogenen[All Fields] OR psychogenenic[All Fields] OR psychogeneous[All Fields] OR
psychogener[All Fields] OR psychogenes[All Fields] OR psychogenese[All Fields] OR
psychogenesis[All Fields] OR psychogenetic[All Fields] OR psychogenetical[All Fields] OR
psychogenetically[All Fields] OR psychogenetics[All Fields] OR psychogenetique[All Fields]
OR psychogenetiques[All Fields] OR psychogenetische[All Fields] OR
psychogenetycznym[All Fields] OR psychogenia[All Fields] OR psychogenias[All Fields] OR

psychogeniatrie[All Fields] OR psychogeniatrie'[All Fields] OR psychogenic[All Fields] OR psychogenic'[All Fields] OR psychogenica[All Fields] OR psychogenical[All Fields] OR psychogenically[All Fields] OR psychogenicity[All Fields] OR psychogenicor[All Fields] OR psychogenics[All Fields] OR psychogenie[All Fields] OR psychogeniecharakterliche[All Fields] OR psychogenien[All Fields] OR psychogenies[All Fields] OR psychogenique[All Fields] OR psychogeniques[All Fields] OR psychogenitat[All Fields] OR psychogenitically[All Fields] OR psychogenna[All Fields] OR psychogenne[All Fields] OR psychogennego[All Fields] OR psychogenneho[All Fields] OR psychogennej[All Fields] OR psychogenni[All Fields] OR psychogennich[All Fields] OR psychogenie[All Fields] OR psychogennih[All Fields] OR psychogennim[All Fields] OR psychogennimi[All Fields] OR psychogeny[All Fields] OR psychogennych[All Fields] OR psychogennym[All Fields] OR psychogennymi[All Fields] OR psychogenomic[All Fields] OR psychogenomics[All Fields] OR psychogenous[All Fields] OR psychogenously[All Fields] OR psychogenuously[All Fields] OR psychogeny[All Fields])) OR (bouffee[All Fields] AND delirante[All Fields])) OR ("schizophrenia, disorganized"[MeSH Terms] OR ("schizophrenia"[All Fields] AND "disorganized"[All Fields]) OR "disorganized schizophrenia"[All Fields] OR "hebephrenia"[All Fields])) OR (schizophren[All Fields] OR schizophrena[All Fields] OR schizophrenc[All Fields] OR schizophrenci[All Fields] OR schizophrencis[All Fields] OR schizophrencs[All Fields] OR schizophrene[All Fields] OR schizophrene's[All Fields] OR schizophrenek[All Fields] OR schizophrenem[All Fields] OR schizophrenen[All Fields] OR schizophrenengruppe[All Fields] OR schizophrenenproblem[All Fields] OR schizophrener[All Fields] OR schizophrenes[All Fields] OR schizophrenese[All Fields] OR schizophreni[All Fields] OR schizophreni'as[All Fields] OR schizophrenia[All Fields] OR schizophrenia'[All Fields] OR schizophrenia's[All Fields] OR schizophrenia1[All Fields] OR schizophreniab[All Fields] OR schizophreniac[All Fields] OR schizophreniaclinicaltrials[All Fields] OR schizophreniacs[All Fields] OR schizophreniaes[All Fields] OR schizophreniaforum[All Fields] OR schizophreniagene[All Fields] OR schizophreniagroup[All Fields] OR schizophreniai[All Fields] OR schizophreniak[All Fields] OR schizophrenial[All Fields] OR schizophrenialike[All Fields] OR schizophreniaor[All Fields] OR schizophreniara[All Fields] OR schizophreniaresearch[All Fields] OR schizophreniaresearchforum[All Fields] OR schizophreniarol[All Fields] OR schizophrenias[All Fields] OR schizophrenias'[All Fields] OR schizophreniaspectrum[All Fields] OR schizophreniay[All Fields] OR schizophrenic[All Fields] OR schizophrenic'[All Fields] OR schizophrenic's[All Fields] OR schizophrenical[All Fields] OR schizophrenically[All Fields] OR schizophrenices[All Fields] OR schizophrenicity[All Fields] OR schizophreniclike[All Fields] OR schizophrenicpatients[All Fields] OR schizophrenics[All Fields] OR schizophrenics'[All Fields] OR schizophrenicss[All Fields] OR schizophrenie[All Fields] OR schizophrenie'[All Fields] OR schizophrenieaehnlichen[All Fields] OR schizophrenieaehnliche[All Fields] OR schizophrenieaehnlichen[All Fields] OR schizophrenieaehnlicher[All Fields] OR schizophrenieartige[All Fields] OR

schizophrenieartigen[All Fields] OR schizophrenieartiger[All Fields] OR
schizophreniebegriffs[All Fields] OR schizophreniebegriff[All Fields] OR
schizophreniebegriffes[All Fields] OR schizophreniebegriffs[All Fields] OR
schizophreniebehandlung[All Fields] OR schizophreniebehandlungen[All Fields] OR
schizophreniediagnose[All Fields] OR schizophrenieerkrankten[All Fields] OR
schizophrenieerkrankter[All Fields] OR schizophreniefalle[All Fields] OR
schizophreniefallen[All Fields] OR schizophrenieform[All Fields] OR schizophrenieforme[All
Fields] OR schizophreniformen[All Fields] OR schizophreniformer[All Fields] OR
schizophrenieforschung[All Fields] OR schizophreniefrage[All Fields] OR
schizophreniegenes[All Fields] OR schizophreniekonzepte[All Fields] OR
schizophreniekonzepten[All Fields] OR schizophreniekrank[All Fields] OR
schizophreniekranken[All Fields] OR schizophreniekranker[All Fields] OR
schizophreniekreises[All Fields] OR schizophrenielehre[All Fields] OR schizophrenien[All
Fields] OR schizophrenienahe[All Fields] OR schizophreniepatienten[All Fields] OR
schizophreniepatientinnen[All Fields] OR schizophrenieproblem[All Fields] OR
schizophrenieproblems[All Fields] OR schizophrenierisiko[All Fields] OR
schizophrenieritoriality[All Fields] OR schizophrenies[All Fields] OR
schizophreniespektrums[All Fields] OR schizophreniespezifitat[All Fields] OR
schizophreniestudie[All Fields] OR schizophreniesymptomen[All Fields] OR
schizophrenietheorie[All Fields] OR schizophrenietherapie[All Fields] OR
schizophrenietypologie[All Fields] OR schizophrenieverlauf[All Fields] OR
schizophrenieverlaufe[All Fields] OR schizophrenieverlaufs[All Fields] OR
schizophrenieverstandnis[All Fields] OR schizophrenifallen[All Fields] OR
schizophreniforms[All Fields] OR schizophreniform[All Fields] OR schizophreniform'[All
Fields] OR schizophreniforme[All Fields] OR schizophreniformen[All Fields] OR
schizophreniformes[All Fields] OR schizophreniformic[All Fields] OR schizophreniforms[All
Fields] OR schizophrenigenesis[All Fields] OR schizophreniics[All Fields] OR
schizophrenikern[All Fields] OR schizophrenine[All Fields] OR schizophreniologists[All
Fields] OR schizophreniphorm[All Fields] OR schizophrenique[All Fields] OR
schizophreniques[All Fields] OR schizophrenis[All Fields] OR schizophrenisation[All Fields]
OR schizophrenism[All Fields] OR schizophreniucs[All Fields] OR schizophrenix[All Fields]
OR schizophrenix's[All Fields] OR schizophreniz[All Fields] OR schizophrenization[All Fields]
OR schizophrenized[All Fields] OR schizophrenieforschung[All Fields] OR schizophrenlcs[All
Fields] OR schizophrenle[All Fields] OR schizophreno[All Fields] OR schizophrenoform[All
Fields] OR schizophrenogenesis[All Fields] OR schizophrenogenic[All Fields] OR
schizophrenoid[All Fields] OR schizophrenomimetic[All Fields] OR schizophrenomimetics[All
Fields] OR schizophrenosimilar[All Fields] OR schizophrenos[All Fields] OR schizophrenuc[All
Fields] OR schizophreny[All Fields])) OR (schizoaffektive[All Fields] OR schizoaffektiv[All
Fields] OR schizoaffektifs[All Fields] OR schizoaffektive[All Fields] OR schizoaffektive'[All

Fields] OR schizoaffectively[All Fields] OR schizoaffectives[All Fields] OR
schizoaffectives'[All Fields] OR schizoaffectivity[All Fields])) OR ("bipolar disorder"[MeSH
Terms] OR ("bipolar"[All Fields] AND "disorder"[All Fields]) OR "bipolar disorder"[All Fields]
OR "manic"[All Fields]) AND ("stupor"[MeSH Terms] OR "stupor"[All Fields])) NOT
comment[Publication Type] NOT case reports[Publication Type] NOT letter[Publication
Type] NOT review[Publication Type]

Appendix II. Results of categorisation of the literature: Complete list of model induction methods used, behavioural outcomes measured, and treatment compounds tested

Results of phases I and II of the project including screening and categorization of entire dataset identified in the search can be viewed here:

Bahor, Zsanett. (2018). Systematic Review of Animal Models of Psychotic Disorders (Version Version I) [Data set]. Zenodo. Accessible at:

<https://doi.org/10.5281/zenodo.1209832>.

Overall frequency lists of methods of model induction, behavioural outcome measures and compounds tested as treatment are available here:

https://drive.google.com/drive/folders/1hMAJboVD8WAGrt_8iRiPs2rkGdzoO_La?usp=sharing

Appendix III. Text mining tool: code used for program and list of expressions

Code

Code for program developed is available here:

https://drive.google.com/open?id=1hMAJboVD8WAGrt_8iRiPs2rkGdzoO_La

Please use the password *TextMiner2018* when unzipping the file.

Simple word phrases

Random allocation to group

randomly assigned, Randomly assigned, assigned randomly, randomly divided, divided randomly, randomly treated, randomly split, randomly determined, random assignment, randomly received, were randomised, were randomized, randomly allocated, allocated randomly, treated in a randomized manner, treated in a randomised manner, randomly selected groups received, randomised to, randomised into, randomised in, randomized to, randomized into, randomized in, random assignment, randomization to, randomisation to, randomly categor, categorised randomly, categorized randomly, randomly subdivided, subdivided randomly, randomly divided, divided randomly

Blinded assessment of outcome

blind to, blindly performed, blinded, blindly, blind manner, blind evaluation, performed blind, blind condition, 'blind', "blind", blind as to, counted blind, blind observer, blind investigator, blind rater, blind experimenter, blind researcher, blind tester, blind quantification, conducted blind, genotype blind, blind coded, blind with respect to, blind method, blind analysis, was unaware, was not aware of, observer unaware of, observer not aware of, observers unaware of, observers not aware of, experimenter unaware of, experimenter not aware of, experimenters unaware of, experimenters not aware of, researcher unaware of, researcher not aware of, researchers unaware of, researchers not aware of, tester unaware of, tester not aware of, testers unaware of, testers not aware of, rater unaware of, rater not aware of, raters unaware of, raters not aware of, person not aware of, person unaware of, investigator unaware of, investigator not aware of, investigators unaware of, investigators not aware of, operator unaware of, operator not aware of, operators unaware of, operators not aware of, were unaware of, were not aware of, without

awareness of, unaware of treatment, unaware of the treatment, unaware of the pretreatment, not aware of treatment, not aware of the treatment, was kept unaware, unaware of genotype, unaware of the genotype, not aware of genotype, not aware of the genotype, kept unaware of, unaware of group, not aware of group, unaware of the group, not aware of the group, unaware of drug, not aware of drug, unaware of the drug, not aware of the drug, unaware of experimental condition, not aware of experimental condition, unaware of the experimental condition, not aware of the experimental condition, naive to the identity of

Sample size calculation

minimum number required to give, required to give statistically valid results, Through a priori calculation, through a priori calculation, minimum sample size, Minimum sample size, planned sample size, calculated a sample size of, Target sample size, target sample size, Sample size of at least, Sample size calculation, Sample size determination, Sample size estimation, Sample size was calculated, Sample size estimate, Sample size consideration, Sample size would be sufficient, sample size of at least, sample size calculation, sample size determination, sample size estimation, sample size was calculated, sample size estimate, sample size consideration, sample size would be sufficient, Power calculation, Power analysis, Power estimation, Power of at least, Power of the study was, power calculation, power analysis, power estimation, power of at least, power of the study was, adequate to detect a, To detect differences of, To detect the treatment effect, To detect a treatment effect, To detect a treatment interaction, To detect statistical differences, To detect the expected difference, To detect the predetermined effect, To detect a mean difference, To detect a difference, To detect a similar treatment effect, To detect a significant change, to detect differences of, to detect the treatment effect, to detect a treatment effect, to detect a treatment interaction, to detect statistical differences, to detect the expected difference, to detect the predetermined effect, to detect a mean difference, to detect a difference, to detect a similar treatment effect, to detect a significant change, A power of, a power of, To insure sufficient power, to insure sufficient power, Sufficient statistical power, sufficient statistical power, would be needed if the null hypothesis, to achieve statistical significance, To achieve statistical significance, to estimate the sample size, To estimate the sample size, was powered at

(mice|rats|animals|subjects|patients) were used|\% chance of|\% to detect
a|minimum clinically worthwhile effect|power of more than|\% to reject the null
hypothesis|effectively powered|power and statistical analysis|are required per
group|per group were required|minimum number required to give|required to give
statistically valid results|through a priori calculation)/i